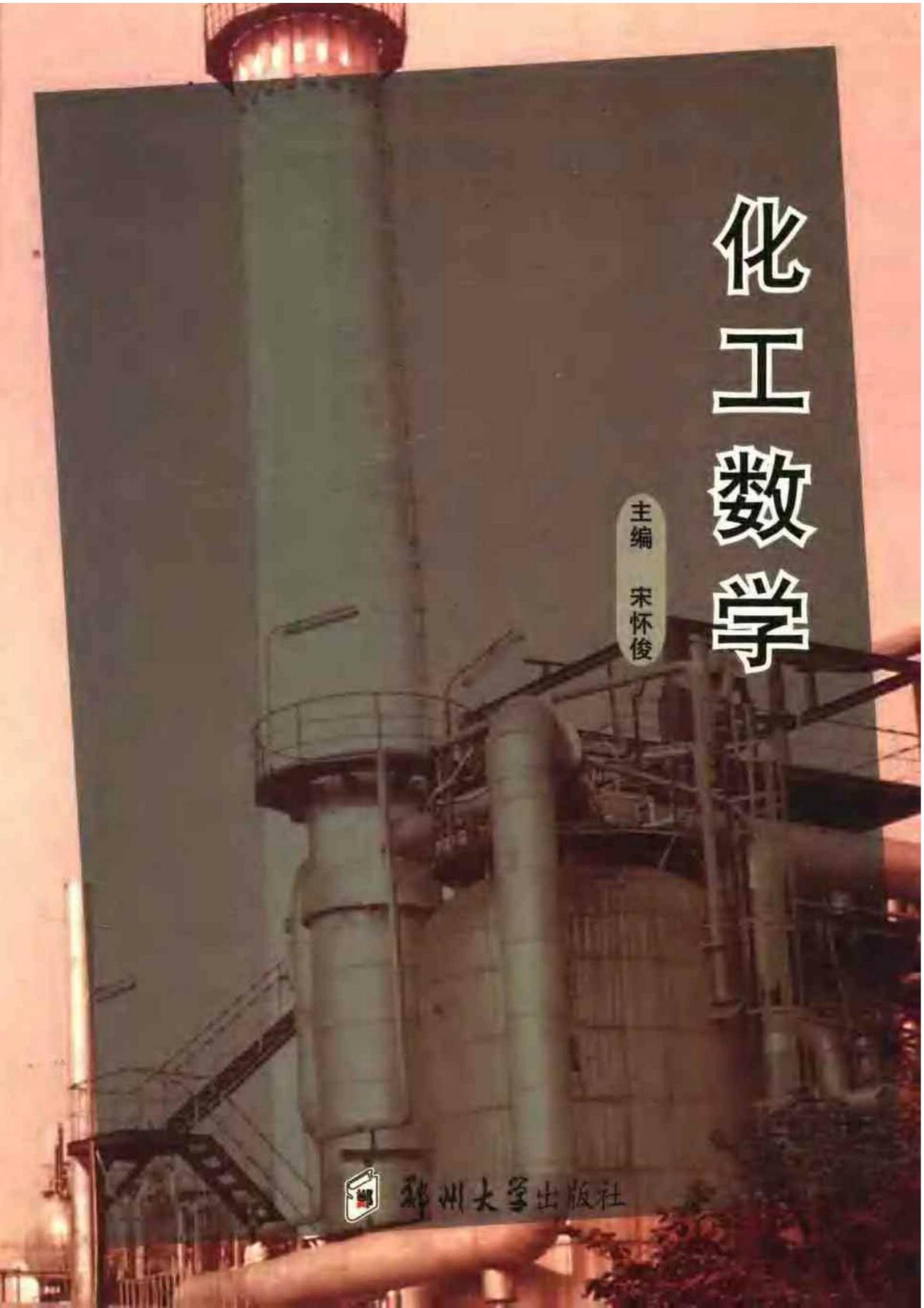
化工数学

主编 宋怀俊



郑州大学出版社

○选题策划　杨秦予
○责任编辑　刘　开
○责任校对　高军晓
○封面设计　高　云
○版式设计　小　羽

ISBN 7-81106-328-X

9 787811 063288 >

ISBN 7-81106-328-X/T · 18

定价：26.00元

化工数学

主编　宋怀俊

郑州大学出版社

图书在版编目(CIP)数据

化工数学/宋怀俊编著. —郑州:郑州大学出版社,
2006.7
ISBN 7-81106-328-X

Ⅰ.化… Ⅱ.宋… Ⅲ.化学工业-应用数学
Ⅳ.TQ011

中国版本图书馆 CIP 数据核字(2006)第 008997 号

郑州大学出版社出版发行
郑州市大学路 40 号　　邮政编码:450052
出版人:邓世平　　发行部电话:0371-66966070
全国新华书店经销
河南东方制图印刷有限公司
开本:710 mm×1 010 mm　　1/16
印张:13
字数:287 千字　　印数:1~4 100
版次:2006 年 7 月第 1 版　　印次:2006 年 7 月第 1 次印刷

书号:ISBN 7-81106-328-X/T·18　　定价:26.00 元

作者名单

主　编　宋怀俊

编　委　宋怀俊　任保增　韩绿霞　李伟然

内容提要

本书在工科高等数学、计算机基础、物理化学、化工原理等课程的基础上，主要介绍了化工过程中遇到的数学问题。全书共分八章，除第一章绪论外，其他七章均为化工计算中的常用方法，分别为：误差分析、非线性方程的数值解法、线性代数方程组的解法、函数的多项式插值、函数的多项式逼近、数值微分与数值积分、常微分方程的数值解法。

书中内容精练，叙述通俗易懂，便于自学。为加强练习，对所讨论的计算方法，只给出理论根据和计算步骤，省去了相应的计算机程序和框图，并在每章后配有习题。本书适合于化工专业本科生、研究生以及工程技术人员参阅。

前　言

现代化工过程是一个由不同类型的化工单元组成的庞大而又复杂的工业系统。在化工过程的研制开发、化工工艺设计、先进化工装置的消化以及现行工业生产的技术改造等方面，都不可避免地要涉及数学计算，而与计算机密切相关的数值方法是其主要内容。同时现代化工过程发展的一个重要标志即是模型化。因此，如何运用现代数学计算工具解决化学工业中的实际问题，实现数学在化学工业中的应用和化工过程的数学模型化，完成数学与化工技术的完美结合，应该是未来化工类工程技术人员必备的知识。

本书根据化工类专业以及相关专业的教学内容和工程实际的需要，强调数学在化工过程中的实际应用，主要介绍了化工过程中遇到的数学问题。全书共分八章，除绪论外，均为化工计算中的常用方法，包括误差分析、非线性方程和线性代数方程组的计算方法、函数的多项式插值、函数的多项式逼近、数值微分、数值积分以及常微分方程的数值解法等。

本书第一、六、七、八章由宋怀俊编写；第二章由任保增编写；第三章由李伟然编写；第四、五章由韩绿霞编写。全书由宋怀俊负责统稿、定稿，由雒廷亮主审。本书在编写过程中得到了郑州大学出版社和郑州大学教务处的大力支持，在此向他们表示衷心的感谢！由于编者水平有限，且编写时间仓促，书中错误之处在所难免，还望广大读者不吝批评、指正，以便使之更加完善。

编　者

2005 年 6 月

目　　录

第一章 绪 论

一、化工过程的数学表示

化学工程是研究大规模地改变物料的化学组成及物理性质的工程技术学科，它研究的内容，不但包括具有化学变化的过程，而且还包括分离混合物为较纯净的不同组分以及改变物理状态和性质的各种过程。化学工程在20世纪初几乎纯属经验，主题是如何利用实验室实验的结果来设计单元设备的容量，即所谓的单元操作时期。到20世纪50年代中期，随着生产的发展，化学工程逐步向技术科学发展，即要求明了化工过程中质量、能量和动量传递的基本现象，这就需要使用数学方法对这些基本现象进行描述。目前，电子计算机的普遍应用，又推动了化学工程进一步走向数学模型化。

所谓数学模型，从广义上讲，就是所考虑的化工过程中某些变量间关系的总称，这种关系，不管是静态的或是动态的，通常都可用公式、表格或图形来表述，这些公式、表格或图形就称为化工过程的数学模型。然而，由于电子计算机的使用，所有表格或图形均可回归成数学表达式，因此，数学模型通常又是指过程的解析表达式。从狭义上讲，数学模型必须是由数学解析表达式构成。从这方面来说，数学模型是可以描写化工过程特性的方程式或方程组。

用数学模型从本质上描述某一化工过程，这不仅需要广泛的数学知识，还必须具有足够的化工专业知识以及对化工过程系统的全面了解。首先要根据过程中化学或物理实际现象及真实过程的物理概念，经过适当的假设和简化建立过程的物理模型；然后再经过必要的归纳和数学推导建立数学模型；最后应用数学方法求解这些数学模型，再应用这些数学解来定量地说明实际过程，从而达到定量分析和预测实际过程的目的。这种模型化的研究方法在化工过程的开发、设备设计及操作条件的优化和过程机理的研究等方面越来越显示出强有力的作用。

二、化工数值方法的意义

在用数学方法解决化工问题的时候，数学模型的建立是化学工程学科的任务，而数学模型的求解则是数学的内容。数学分析及代数学提供了各类数学问题的解析解法，但能够给出精确的解析解的数学问题是很有限的。这样就不得不求助于数值方法来提供各类数学问题的数值解。数值解虽然是近似而且离散的，但它可用于处理极其广泛的工程数学上的多种问题。

由于化工领域涉及面广、过程复杂，它提出了相当多的复杂数学模型及数学问题，涉及到许多数学分支知识，包括线性和非线性代数方程、微分方程、数值微分和数值积分等。对于这类复杂的数学模型，经典的数学解析法已无能为力，必须借助于数值方法，应用计算机求解。因此，数值方法在化工领域内占有极其重要的位置，是化学工程专业技术人员不可缺少的基础知识，也是现代化工技术发展的促进因素。

随着计算机技术的发展和工程问题的需要，目前已涌现出大量功能强、精度高的计算机软件，这为工程技术人员解决实际工程问题提供了相当方便的条件。但只有掌握了数值方法，才能合理地选择和有效地应用这些软件解决实际计算问题。因为在使用这些软件去解决具体的工程问题时，通常会遇到如下几个问题：

(1)数学模型是否准确地反映了实际化学和物理过程。

(2)选用的数值方法是否合适，方法的误差是否超过工程问题允许的误差。

(3)选用的软件提供的程序的实际使用条件是否恰当，在解决具体工程问题时应作哪些修改或调整等等。

工程计算中，这些问题必须搞清楚，否则就可能发生偏差，甚至导致计算失败。

本书是本着为化学工程专业技术人员提供必要的数值方法的基础知识而编写的。全书在论述上着重于各种数值方法的介绍和应用，尽量避免过多的数学证明与推导，使读者能较快地掌握和应用这些方法。对于某些数学方法的讨论，为便于读者理解，仅简要地证明了定理的正确性。

第二章　误差分析

在数值计算中，一般来说，求得的数学模型的解和参加运算的数值都是近似的，也就是说，它们带有一定的误差。例如用观测实验的方法得到的参加运算的数据，它们必然带有误差；也有一些数据，如 π、e、$\sqrt{2}$、$\sqrt{3}$等这些无理数，只能用有理数近似地表示，也会产生误差。这些原始数据所带的误差随着计算过程的演变，往往会产生积累与扩散，给计算的结果造成严重的影响。在许多工程计算中，虽然不必追求真解，但却要求了解近似解的误差范围。

因此，研究误差产生的原因、误差在计算过程中的积累与传播情况，以及如何尽可能地减少误差以保证运算结果的可靠性，是每一个从事数值计算的工作者应当首先考虑的问题。本章的目的就是对这些问题进行概括的说明。

第一节　误差对计算结果的影响及其分类

一、误差对计算结果的影响

现通过一个例子来说明近似数的误差，以及在计算过程中误差的积累与传播对计算结果的影响。

例 2-1　设一半径为 r 的球与两个相互垂直的平面相切，另有一底面半径为 R 的圆柱与球及两平面都相切，如图 2-1 所示，试求这个球的体积。（R 已知）

用初等几何的方法可得

$$r = R\left(\frac{\sqrt{2}-1}{\sqrt{2}+1}\right)$$

从而球的体积为

$$V = \frac{4}{3}\pi R^3\left(\frac{\sqrt{2}-1}{\sqrt{2}+1}\right)^3$$

下面只就括号部分的立方进行计算。

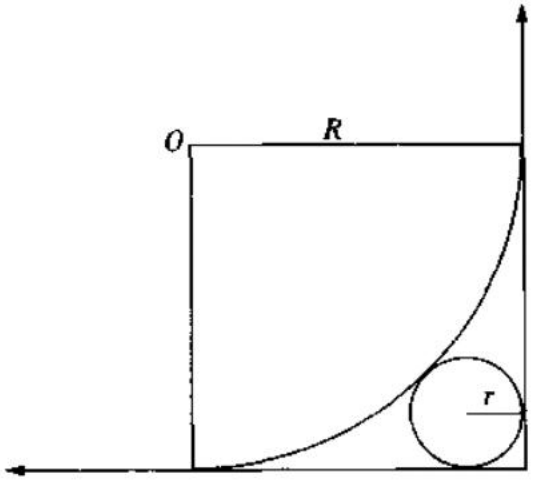


图 2-1

设

$$X=\left(\frac{\sqrt{2}-1}{\sqrt{2}+1}\right)^3$$

这个式子可以用以下 6 种形式来表示：

$$X=(\sqrt{2}-1)^6=(3-2\sqrt{2})^3=99-70\sqrt{2}$$

$$X=\left(\frac{1}{\sqrt{2}+1}\right)^6=\left(\frac{1}{3+2\sqrt{2}}\right)^3=\frac{1}{99+70\sqrt{2}}$$

由于$\sqrt{2}$的真值是 1.414 213…，若取$\sqrt{2}$的两个近似值为：$\sqrt{2}=1.4$ 和$\sqrt{2}=1.41$，分别按以上 6 个公式计算，结果如表 2-1 所示。

表 2-1

$\sqrt{2}$	1.4	1.41
$(\sqrt{2}-1)^6$	0.004 096	0.004 750
$(3-2\sqrt{2})^3$	0.008 000	0.005 832
$99-70\sqrt{2}$	1.000 000	0.300 000
$\left(\frac{1}{2\sqrt{2}+1}\right)^6$	0.005 233	0.005 104
$\left(\frac{1}{3+2\sqrt{2}}\right)^3$	0.005 125	0.005 073
$\frac{1}{90+70\sqrt{2}}$	0.005 076	0.005 058

从计算的结果可以看出，按不同公式，取不同的近似值进行计算，所得结果是不一样的，有的相差甚远，究竟取哪一个接近于真值？这就是我们要研究的问题。

二、误差的分类

由例 2－1 可见，原始数据与计算结果之间通常或多或少地存在一定的误差，这些误差的来源大体上可分为以下 5 个方面。

1. 模型误差

模型误差指的是数学方法的描述与实际物理现象之间的误差。根据实际工程问题建立数学模型时，必须经过某种程度的简化和归纳，即所谓“理想化”处理。这种处理一方面是为了使具体问题抽象化，使模型具有普遍的适应性；另一方面也是为了使模型简化，以便于进行数学处理。这就造成了模型与实际问题之间的误差。例如固定床反应器的拟均相模型、多元精馏塔的平衡级模型、伴随有化学反应的吸收塔模型等等，都作过种种理想化处理。但应明确，这类误差不是数值方法的某种算法所造成的，而是算法所采用的模型本身带来的。

2. 观测误差

在数值计算中，相当多数量的原始数据是用观测、试验的方法得到的。在得到这些数据的过程中，由于受到工作人员的生理条件、实验手段以及技术水平的限制，所得数据不可能是完全精确的，这种实验观测结果与真值之间的误差称为试验观测误差。

3. 截断误差

在工程问题计算中，并不是所有的问题都能通过对数学模型的直接计算而取得结果的。常常会遇到超越运算，这就要求用极限和无穷过程来表示。但在实际计算时，只能进行有限次运算，因而只能求得其近似值。例如，在数值方法中，常采用收敛级数的前 n 项代替无穷级数，也就是舍去了 n 项以后的项。由此引起的误差称为截断误差。这类误差与数值方法中的算法有关，在算法选择和进行具体运算时，应注意对这类误差的分析。实际上，常用截断误差限或截断误差的阶来判定某种算法的优劣。

4. 舍入误差

实际计算中，无论是由观测得来的数据或是由计算得来的数据，其位数总应该是有限位。因此就需要对数据进行舍入，得到一定位数的近似值。这样产生的误差称为舍入误差。数据舍入的方法有多种，最常用的即熟知的四舍五入法则。

5. 过失误差

这种误差的产生是由于计算人员的粗枝大叶或疏忽大意所造成的，其结果是难以预料的，原因往往不易找到。如听错、读错、写错以及建立的模型和选择的数值方法不合实际或不科学等等。应指出的是，过失误差对于一种严肃的科学计算来说是不允许的。这就要求计算工作者始终保持严谨的工作态度，以利于消除过失误差。

以上概述了误差的各种来源。在数值计算乃至程序设计中，主要关心与研究的是截断误差与舍入误差对计算结果的影响。

第二节　误差的表示和有效数字

在数值计算中，无论是原始数据，还是计算结果，一般说来都是近似值，也即它们与真值之间存在有一定的误差。衡量近似值与真值的接近程度无疑是人们所关心的问题。通常把近似值与真值的接近程度称为近似值的准确度。为了有效地衡量一个近似值的准确度高低，现引入误差的表示形式和近似值有效数字的概念。

一、误差表示

在数值计算或测量中，常用的误差表示方法有绝对误差、相对误差、标准误差等等。

1. 绝对误差

设真值 x 的近似值为 x^*，则称 $x-x^*$ 为 x 的绝对误差，记为：

$$e=x-x^*$$

在一般情况下，真值是不能得到的，因此，也就无法得到绝对误差。但可根据近似值本身的性质以及取舍法则来估算出这个绝对误差的范围，即找到一个尽可能小的正数 ε，使得绝对误差的绝对值不超过这个正数，其中，ε 称为绝对误差界。即

$$|e|=|x-x^*|\leqslant\varepsilon$$

在工程技术上，绝对误差界常以正负偏差的形式出现。真值、近似值及绝对误差界之间的关系表示为：

$$x=x^*\pm\varepsilon$$

例如，$G=200\ \text{kg}\pm0.1\ \text{kg}$，表示近似值 $G=200\ \text{kg}$，其绝对误差界是 $0.1\ \text{kg}$。从绝对误差与绝对误差界的意义可以知道它们是有单位的。

2. 相对误差

近似值 x^* 的相对误差为：

$$e_r=\frac{e}{x^*}=\frac{x-x^*}{x^*}$$

相对误差不仅可以表示近似值 x^* 的误差大小，还可以完善地表示其准确程度，比绝对误差的概念更为完善。例如，假设称量 $1\ 000\ \text{kg}$ 与 $100\ \text{kg}$ 的两种物体时，所得绝对误差均为 $1\ \text{kg}$，那么，应当认为对 $1\ 000\ \text{kg}$ 物体的称量准确度比对 $100\ \text{kg}$ 物体的称量准确度要高。故相对误差比绝对误差更能准确地表示一个近似值的准确度。

与绝对误差一样，相对误差也摆脱不了真值 x 的影响。通常是找一个尽可能小的正数 ε_r，使得相对误差的绝对值最大时也不超过它，即

$$|e_r|=\left|\frac{e}{x^*}\right|=\left|\frac{x-x^*}{x^*}\right|\leqslant\varepsilon_r=\left|\frac{\varepsilon}{x^*}\right|$$

称这个 ε_r 为近似值 x^* 的相对误差界。

真值、近似值及相对误差界之间的关系可表示为：

$$x = x^*(1 \pm \varepsilon_r)$$

相对误差与相对误差界为无名数，常用百分数来表示。

3. 标准误差

标准误差(σ)的定义式为：

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (n \text{ 为无限大，即测定次数 } n \text{ 为无穷多})$$

或

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (n \text{ 为有限次测定})$$

其中，$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$。

标准误差也称为均方根误差，它是工程计算中常用的表示误差的方法。这种误差的表示方法虽不直观，但它消除了正负偏差相互抵消的可能性，它不但与一组测定值中每个数据有关，而且对其中较大误差或较小误差的敏感性很强，能较明显地反映出较大的个别误差。实验愈精确，其标准误差愈小。

二、有效数字

1. 有效数字的定义与一般形式

(1)定义：若近似值 x^* 的绝对误差不超过某位数字的半个单位，那么从该位数字到 x^* 最左边的那个非零数字(设共有 n 位)都称为 x^* 的有效数字。

例如，$|e - 2.718| < 0.0005 = \frac{1}{2} \times 0.001$，则 e 的近似值 2.718 有 4 位有效数字。

再如，若近似值 123.45 与 876.000 是用四舍五入法则得来的，那么前者有 5 位有效数字，后者有 6 位有效数字。

(2)一般形式：任何一个具有 n 位有效数字的近似值 x^* 总可以表示成下列一般形式

$$\begin{aligned} x^* &= \pm 10^m[a_1 + a_2 \times 10^{-1} + \cdots + a_n \times 10^{-(n-1)}] \\ &= \pm 10^{m+1}(a_1 \times 10^{-1} + a_2 \times 10^{-2} + \cdots + a_n \times 10^{-n}) \end{aligned}$$

其中，n 为有效数字位数；m 为任何整数；a_i 为 0 ~ 9 中的整数，且 $a_1 \neq 0$。

同时，显然有

$$|x - x^*| \leqslant \frac{1}{2} \times 10^m \times 10^{-(n-1)} = \frac{1}{2} \times 10^{m-n+1}$$

这就说明了有效数字与绝对误差之间的关系。

例如重力加速度 g，若四舍五入到小数点后两位，通常得 $g = 9.80\ m/s^2$，即

$$|g - 9.80| \leqslant \frac{1}{2} \times 10^{-2}$$

因为

$$9.80 = 10^{0}(9 + 8 \times 10^{-1} + 0 \times 10^{-2})$$

于是

$$-(n-1) = -2$$

$$n = 3$$

即 9.80 具有 3 位有效数字。

现把 g 的单位改为 km/s^2，则 $g = 0.009\ 80\ km/s^2$。

由于

$$0.009\ 80 = 10^{-3}(9 + 8 \times 10^{-1} + 0 \times 10^{-2})$$

同样

$$-(n-1) = -2$$

$$n = 3$$

这说明 g 仍为 3 位有效数字，事实上这是和有效数字的规定相符合的。

2. 有效数字与相对误差的关系

从上段的分析可以看出，近似值 x^* 的有效位数越多，绝对误差界就越小，反之亦然。那么有效数字与相对误差的关系如何呢？它们之间的关系可以用以下两条定理来说明。

定理 1 若近似值 x^* 具有 n 位有效数字，那么它的相对误差满足

$$|e_r| \leqslant \frac{1}{2a_1} \times 10^{-(n-1)}$$

证明 从具有 n 位有效数字的近似值 x^* 的一般表达式得知

$$|x^*| \geqslant a_1 \times 10^{m}$$

$$|e_r| = \left|\frac{e}{x^*}\right| \leqslant \frac{1}{a_1 \times 10^{m}} \times \frac{1}{2} \times 10^{m-(n-1)} = \frac{1}{2a_1} \times 10^{-(n-1)}$$

由定理 1 可知，有效位数越多，相对误差越小。

定理 2 若近似值 x^* 的相对误差 e_r 满足

$$|e_r| \leqslant \frac{1}{2(a_1+1)} \times 10^{-(n-1)}$$

则 x^* 至少有 n 位有效数字。

证明 由于

$$e = x^* e_r$$

$$|x^*| = |\pm 10^{m}[a_1 + a_2 \times 10^{-1} + \cdots + a_n \times 10^{-(n-1)}]| \leqslant (a_1 + 1) \times 10^{m}$$

则

$$|e| = |x^*| \cdot |e_r| \leqslant (a_1 + 1) \times 10^{m} \times \frac{1}{2(a_1+1)} \times 10^{-(n-1)}$$

即

$$|e| \leqslant \frac{1}{2} \times 10^{m-(n-1)}$$

故 x^* 至少有 n 位有效数字。

从以上两个定理可以看出，有效数字是近似值精度的重要标志。

例 2-2 要使 $\sqrt{30}$ 的近似值的相对误差不超过 0.1%，应取几位有效数字？

解 由定理 1 可知

$$|e_r| \leqslant \frac{1}{2a_1} \times 10^{-(n-1)}$$

因为 $a_1 = 5$，令

$$\frac{1}{2 \times 5} \times 10^{-(n-1)} \leqslant 0.1\%$$

则 $n \geqslant 3$，即应取 3 位有效数字。

按定理 2 应有

$$\frac{1}{2(a_1+1)} \times 10^{-(n-1)} \leqslant 0.1\%$$

$$\frac{1}{2(5+1)} \times 10^{-(n-1)} \leqslant 0.1\%$$

同样取 $n = 3$ 即可满足要求。

第三节 算术运算中的误差积累与传播

一、一般函数的误差计算

从微分学知道，函数的微分是函数增量的线性主部，它们之间相差一个高阶的无穷小量。因而用函数的微分来代替函数的增量不仅是可能的，而且往往可以简化计算。这里所说的增量，指的就是函数的绝对误差。

设多元函数 $F = f(x_1, x_2, \cdots, x_n)$，自变量 $x_1, x_2, \cdots, x_n$ 的近似值分别为 $x_1^*, x_2^*, \cdots, x_n^*$，函数 F 的近似值为 $F^* = f(x_1^*, x_2^*, \cdots, x_n^*)$，则函数近似值的绝对误差为（各自变量的绝对误差分别记为 $e_1, e_2, \cdots, e_n$）

$$\begin{aligned} e(F^*) = F - F^* &= f(x_1, x_2, \cdots, x_n) - f(x_1^*, x_2^*, \cdots, x_n^*) \\ &= f(x_1^* + e_1, x_2^* + e_2, \cdots, x_n^* + e_n) - f(x_1^*, x_2^*, \cdots, x_n^*) \end{aligned}$$

由微分与增量的关系可知

$$e(F^*) = \mathrm{d}f(x_1^*, x_2^*, \cdots, x_n^*) + \delta$$

即

$$e(F^*) = \frac{\partial f}{\partial x_1^*} e_1 + \frac{\partial f}{\partial x_2^*} e_2 + \cdots + \frac{\partial f}{\partial x_n^*} e_n + \delta$$

当略去高阶无穷小量 δ 时可得

$$e(F^*) = \frac{\partial f}{\partial x_1^*} e_1 + \frac{\partial f}{\partial x_2^*} e_2 + \cdots + \frac{\partial f}{\partial x_n^*} e_n$$

在上述公式中若以绝对误差界来代替绝对误差，并将各偏导数取绝对值，则得函数的绝对误差界：

$$\varepsilon(F^*) \leqslant \left|\frac{\partial f}{\partial x_1^*}\right|\varepsilon_1 + \left|\frac{\partial f}{\partial x_2^*}\right|\varepsilon_2 + \cdots + \left|\frac{\partial f}{\partial x_n^*}\right|\varepsilon_n$$

这就是函数误差法则的一般公式，由此可以推导出各种算术运算的误差法则。

二、算术运算的误差法则

1. 近似值代数和的误差法则

设函数 $f(a,b) = a \pm b$，则

$$\left|\frac{\partial f}{\partial a}\right| = 1, \quad \left|\frac{\partial f}{\partial b}\right| = 1$$

于是

$$\varepsilon(a \pm b) \leqslant \varepsilon(a) + \varepsilon(b)$$

即两个近似值代数和与差的绝对误差界不超过各自绝对误差界的和。

例 2-3 已知 $\frac{1}{3}, \frac{1}{7}, \frac{1}{11}, \frac{1}{13}, \frac{1}{15}, \frac{1}{17}$ 的近似值分别为 0.333，0.143，0.091，0.077，0.067，0.059，试求其和并确定和有几位有效数字。

解 $S = 0.333 + 0.143 + 0.091 + 0.077 + 0.067 + 0.059 = 0.770$

和的绝对误差界为

$$\varepsilon(S) = 0.000\,5 \times 6 = 0.003 < \frac{1}{2} \times 0.01$$

则和有两位有效数字，$S = 0.77$。

例 2-4 已知近似值 $a = 124.3, b = 15.78, c = 0.457$，试求它们的和并确定和的有效数字。

解 $S = 124.3 + 15.78 + 0.457 = 140.537$

绝对误差界为

$$\varepsilon(S) = 0.05 + 0.005 + 0.000\,5 = 0.055\,5 < \frac{1}{2} \times 1$$

所以和有 3 位有效数字，舍入后得 $S = 141$。

由例 2-4 可看出，在加法运算中由于小数点要对齐，所以，尽管某些近似数小数点后边的有效位很多，但和的有效位也不会提高。有效的办法是事先对原始数据进行舍入，使小数点之后保留同样的有效位，然后再进行求和运算。

在例 2-4 中可取 $a = 124.3, b = 15.8, c = 0.5$，则

$$S = 124.3 + 15.8 + 0.5 = 140.6$$

$$\varepsilon(s) = 0.05 \times 3 = 0.15 < \frac{1}{2} \times 1$$

即只能有 3 位有效数字，舍入后得 $S = 141$。

有了代数和与差绝对误差界判别法则之后，立即可以得出代数和与差的相对误差界的判别法则：

$$\varepsilon_r(a+b)=\frac{\varepsilon(a\pm b)}{|a\pm b|}\leqslant\frac{\varepsilon(a)+\varepsilon(b)}{|a\pm b|}$$

当 a、b 同号时，和的相对误差界有更简单的估计式：

$$\varepsilon_r(a+b)=\max\left\{\frac{\varepsilon(a)}{|a|},\frac{\varepsilon(b)}{|b|}\right\}$$

证明 假定 $\dfrac{\varepsilon(a)}{|a|}\geqslant\dfrac{\varepsilon(b)}{|b|}$，即 $|b|\varepsilon(a)\geqslant|a|\varepsilon(b)$。

于是

$$|b|\varepsilon(a)+|a|\varepsilon(a)\geqslant|a|\varepsilon(b)+|a|\varepsilon(a)$$

$$\frac{|b|\varepsilon(a)+|a|\varepsilon(a)}{|a|(|a|+|b|)}\geqslant\frac{|a|\varepsilon(b)+|a|\varepsilon(a)}{|a|(|a|+|b|)}$$

则

$$\frac{\varepsilon(a)}{|a|}\geqslant\frac{\varepsilon(a)+\varepsilon(b)}{|a|+|b|}=\frac{\varepsilon(a)+\varepsilon(b)}{|a+b|}$$

2. 近似值乘积的误差法则

设函数 $f(a,b)=a\cdot b$，则 $\left|\dfrac{\partial f}{\partial a}\right|=|b|$，$\left|\dfrac{\partial f}{\partial b}\right|=|a|$。

于是

$$\varepsilon(a\cdot b)\leqslant|a|\varepsilon(b)+|b|\varepsilon(a)$$

上式两边同除以 $|a\cdot b|$ 即得相对误差界估计式

$$\varepsilon_r(a\cdot b)=\frac{\varepsilon(a\cdot b)}{|a\cdot b|}\leqslant\frac{\varepsilon(b)}{|b|}+\frac{\varepsilon(a)}{|a|}=\varepsilon_r(b)+\varepsilon_r(a)$$

即两近似值积的相对误差界不超过两近似值相对误差界的和。

例 2-5 求 32.26×2.13 的积及其有效数字。

解 先求乘积，再决定有效数字。

$$32.26\times2.13=68.713\,8$$

积的绝对误差界

$$\varepsilon(a\cdot b)=(32.26+2.13)\times0.005=0.172$$

则积中只能有两位有效数字，舍入后得积为 69。

例 2-6 设 $a=2.334$，$b=8.21$，求它们的积。

解

$$2.334\times8.21=19.162\,14$$

$$\varepsilon(a\cdot b)=2.334\times0.000\,5+8.21\times0.005=0.041\,05<\frac{1}{2}\times0.1$$

所以积中可以有 3 位有效数字，舍入得 a，b 的积为 19.2。

由以上两个例题可概括出乘积有效数字的一般规律，两个数乘积的有效位数等于因

数中较少的有效位数,或至少等于较少的有效位数减1(如例2-5)。

3. 近似值商的误差法则

设函数$f(a,b)=\frac{a}{b}$,则$\left|\frac{\partial f}{\partial a}\right|=\frac{1}{|b|}$,$\left|\frac{\partial f}{\partial b}\right|=\frac{|a|}{b^2}$。

于是

$$\varepsilon\left(\frac{a}{b}\right)=\frac{1}{b}\varepsilon(a)+\frac{|a|}{b^2}\varepsilon(b)=\frac{|b|\varepsilon(a)+|a|\varepsilon(b)}{b^2}$$

将上式两边同除以$\left|\frac{a}{b}\right|$得商的相对误差界

$$\varepsilon_r\left(\frac{a}{b}\right)=\frac{\varepsilon\left(\frac{a}{b}\right)}{\left|\frac{a}{b}\right|}=\frac{\varepsilon(a)}{|a|}+\frac{\varepsilon(b)}{|b|}=\varepsilon_r(a)+\varepsilon_r(b)$$

即两近似值商的相对误差界不超过各自相对误差界的和。

例2-7 某物体以15.2 m/s的速度作匀速运动,求经过39.4 m路程所需的时间和可能的相对误差界。

解

$$t=\frac{s}{v}=\frac{39.4}{15.2}\approx 2.592\,1(\mathrm{s})$$

$$\varepsilon\left(\frac{s}{v}\right)=\frac{15.2\times 0.05+39.4\times 0.05}{(15.2)^2}\approx 0.011\,8<\frac{1}{2}\times 0.1$$

则商中可有两位有效数字,舍入得$t=2.6(\mathrm{s})$。

$$\varepsilon_r\left(\frac{s}{v}\right)=\varepsilon_r(s)+\varepsilon_r(v)=\frac{0.05}{39.4}+\frac{0.05}{15.2}\approx 0.004\,56=0.456\%$$

与乘积一样,关于商的有效数字也有类似的一般结论:商的有效数字等于运算对象中有效位数较少的那一个,或至少等于其有效位数减1。

4. 近似值乘方的误差法则

设函数$f(a)=a^m$,则$|f'(a)|=|m||a^{m-1}|$。

于是

$$\varepsilon(a^m)=|m||a^{m-1}|\varepsilon(a)$$

而

$$\varepsilon_r(a^m)=\frac{\varepsilon(a^m)}{|a^m|}=|m|\frac{\varepsilon(a)}{|a|}=|m|\varepsilon_r(a)$$

即一个近似值任意次幂的相对误差界不超过底数的相对误差界与指数绝对值的乘积。

特别地,当$m=\frac{1}{2}$时,$\varepsilon_r(\sqrt{a})=\frac{1}{2}\varepsilon_r(a)$;当$m=\frac{1}{3}$时,$\varepsilon_r(a^{\frac{1}{3}})=\frac{1}{3}\varepsilon_r(a)$。

第四节 关于误差分析中的几个问题

在上节中讨论了误差分析的一般方法,但在数值计算中应注意具体问题具体分析,

一定不要拘泥于这些一般方法。下面举出几个数值计算中应注意的问题。

1. 尽量避免两个相近数作减法运算

两相近数相减时，往往产生很大的误差，造成有效数字严重地丢失，从而使计算结果不可用。

产生这种后果的原因是十分明显的，由两数差的相对误差界的估计式可以看出，相对误差界是由一个十分接近于零的数$|a-b|$作除数得到的。这样相对误差就会变得很大，从而严重地影响结果的精度。

例 2-8 假定$(1\ 972)^{\frac{1}{2}}\approx 44.41$，$(1\ 971)^{\frac{1}{2}}\approx 44.40$，试计算

$$X=(1\ 972)^{\frac{1}{2}}-(1\ 971)^{\frac{1}{2}}$$

的值，并求它的绝对误差与相对误差。

解

$$X=44.41-44.40=0.01$$

$$\varepsilon(X)=\varepsilon(44.41)+\varepsilon(44.40)=0.005+0.005=0.01$$

$$\varepsilon_r(X)=\frac{\varepsilon(X)}{X}=\frac{0.01}{0.01}=100\%$$

可见，相对误差是100%。在这种情况下，若用下式来代替减法，则可以收到好得多的效果。

$$X=\frac{1\ 972-1\ 971}{(1\ 972)^{\frac{1}{2}}+(1\ 971)^{\frac{1}{2}}}=\frac{1}{(1\ 972)^{\frac{1}{2}}+(1\ 971)^{\frac{1}{2}}}=\frac{1}{44.41+44.40}\approx 0.011\ 26$$

现按照除法对误差估计，则分母的绝对误差为0.01，分母的相对误差为$\frac{0.01}{88.81}=0.011\%$。由于分子是精确的，相对误差为零，则按照除法误差法则得商的相对误差为0.011%。

为了避免这种情况发生，在计算之前可以对计算公式作适当的数学处理。现举如下几个例子：

(1)计算$\ln x_1-\ln x_2$，当x_1与x_2很接近时可作变换

$$\ln x_1-\ln x_2=\ln\frac{x_1}{x_2}$$

(2)计算$\frac{1-\cos x}{\sin x}$，当x接近于零时，可作变换

$$\frac{1-\cos x}{\sin x}=\frac{1-\cos^2 x}{\sin x(1+\cos x)}=\frac{\sin x}{1+\cos x}$$

(3)计算$\sqrt{1+x}-\sqrt{x}$，当x充分大时，可作变换

$$\sqrt{1+x}-\sqrt{x}=\frac{1}{\sqrt{1+x}+\sqrt{x}}$$

(4)计算$\arctan(x+1)-\arctan x$，当x充分大时可作变换

$$\arctan(x+1)-\arctan x=\arctan\frac{1}{1+x(1+x)}$$

2. 绝对值太小的数不宜作除数

对于除式$\frac{a}{b}$，若b很接近于零，则由商的误差估计法则知道，商的绝对误差可能很大，从而影响近似值的精确度。

3. 注意避免两个绝对值相差很大的数作加减运算

由于计算机的字长总是有限的，而在机器上作加减运算时按数的规格化的指数形式表示的数需要对阶，这时绝对值较小的数往往被较大的数“吃掉”。

例如：$a=10^9, b=1$ 则

$$a+b=0.1\times10^{10}+0.000\,000\,00\,\boxed{0\,1}\times10^{10}=0.1\times10^{10}$$

这种情况往往导致一些计算结果严重地失真，遇到这种情况，可以从算法上重新考虑以避免其发生。

4. 在选择算法时应注意尽量减少运算次数

计算结果的误差往往是通过程序逐步积累而形成的，当然一个好的算法有时也可以使误差不至增加，甚至减少。然而为了避免误差的积累，计算人员应当精心地设计与选择计算方法，以使运算次数尽可能地减少，这样可以有效地减少误差的积累。

例如，计算多项式

$$P_n(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$$

的值，若直接计算，要用$n+(n-1)+\cdots+2+1=\frac{n(n+1)}{2}$次乘法，$n$次加法。如果把原式按秦九韶算法，可以给出下列格式

$$\begin{cases}x=x_0\\P=A(n)\\P=Px+A(i)\quad(i=n-1,n-2,\cdots,2,1,0)\end{cases}$$

其中，x_0为给定的x值；$A(i)$为各个系数；P为从0次到n次的各次多项式。

按照这个格式进行计算，只需n次乘法，n次加法就可求得多项式的值。

5. 增加原始数据的有效位数

在不易采取有效措施改善精度的情况下，在计算机软、硬件允许的范围内尽可能地增加原始数据的有效位数，可提高结果的精度。

习　题

1. 为什么要引入绝对误差、相对误差、有效数字的概念？

2. 为什么要引入绝对误差界、相对误差界的概念？

3. 下列各数均是由四舍五入得到的，试给出它们的绝对误差界、相对误差界及有效位数。

$$a=0.368\,9,\quad b=59.000,\quad c=0.000\,15,\quad d=90.001$$

4. 为使 $\sqrt{150}$ 的近似值的相对误差不超过 0.1%，至少应有几位有效数字？

5. 下列各题在计算时应作何种处理？

(1) $1-\cos 1°$；　　(2) $\ln\left(90-\sqrt{90^2-1}\right)$；

(3) $\int_N^{N+1}\frac{dx}{1+x^2}$，当 N 充分大时；　　(4) e^x-1，当 x 接近于 0 时。

6. 湿空气的比热在 15 ℃时为 0.238 6 kJ/(kg·℃)，且已知其相对误差为 0.5%，试估计有效数字的位数为多少？

7. 含有氯的空气 10.3 L 通过 57.48 g 活性炭，氯被吸附，空气的体积减小到 9.9 L，而活性炭的质量增加到 58.69 g。若氯气的含量：(1)按空气的体积变化计算；(2)按活性炭质量的变化计算。试比较各法的相对误差，并指出采用哪种方法较好？

8. 编一个用秦九韶算法求多项式的值的程序。

第三章　非线性方程的数值解法

本章主要讨论形如$f(x)=0$的方程的求解问题。这类方程通常也称为一元非线性方程。许多工程技术中的问题往往归结为求这样方程的实根。然而这类方程只有在为数不多的情形下可以用代数学的方法即所谓精确的方法找出根的解析表达式，这种方法也称为直接方法。已经证明，对于高于4次的多项式的求根，不存在直接方法。此外，即便是可用直接方法来得到根的表达式，由于原始数据的近似性，也往往只能是得到根的一个近似值。所以研究求非线性方程实根的近似解法也就更具有现实的意义。数值方法求根，首先是确定方程的根的大致范围，即进行根的隔离，从某个初始近似值出发，再按照某种数值过程模式逐次逼近根的准确值，最终得到一个满足精度要求的根的近似值。

第一节　实根的隔离与粗略近似值的获得

一、实根的隔离方法

当方程$f(x)=0$在区间$[a,b]$内有多个根时，必须进行根的隔离，寻找单根区间$[a_i,b_i](i=1,2,\cdots,N)$。常用的实根隔离或初始近似值估计的方法有以下3种。

1. 物理法

数学方程$f(x)=0$来源于实际工程或科学计算问题，因此可根据工程或科学问题中的物理概念确定初值x_0。例如，在计算实际气体的压缩因子时，可将理想气体的压缩因子作为初值，即$Z_0=1$；再如，计算理想溶液的泡点时，可将轻、重组分的沸点T_l和T_m作为根区间端点的值，依照安托万(Antoine)方程

$$T=\frac{B}{A-\ln p}-C$$

由于泡点方程$f(T)=0$在$[T_l,T_m]$内单调连续，则必有$T_0\in[T_l,T_m]$。

物理法估计初值简便而确切，并具有明确的物理概念，但在实际应用上有一定的局限性。例如，气体的$R-k$方程的压缩因子表达式为：

$$f(Z)=Z^3-Z^2+Bp\left(\frac{A}{B}-Bp-1\right)Z-\frac{A}{B}(Bp)^2=0$$

其中，常数 $A=a/(R^2T^{2.5})$，$B=b/(RT)$，A 和 B 均与气体性质有关。

该方程具有 3 个根，最大者 Z_{max} 为气体压缩因子；最小者 Z_{min} 为液体压缩因子；中等者 Z_{mid} 无意义。给定初值 $Z_0=1$ 只能求得 Z_{max}，而求不出 Z_{min}，因此，物理法不能解决所有初值的估计问题。有许多问题的初值估计尚需借助数学法。

2. 数学法

数学上主要是利用零点定理进行根的隔离。

零点定理 若函数 $f(x)$ 在 $[a,b]$ 上连续，并且 $f(a)\cdot f(b)<0$，则至少存在一个数 $\xi\in(a,b)$，使得 $f(\xi)=0$。若同时 $f'(x)$ 在 $[a,b]$ 内存在且保持定号，即 $f(x)$ 在 $[a,b]$ 上单调时，则这样的 ξ 在 (a,b) 内是唯一的。

在应用定理的前半部分的时候，可以把满足条件的区间逐渐细分，例如可以采用逐步加密等分的方法把各个不同的根置于小区间上。特别是对于多项式方程来说，n 次方程的实根数目不超过 n 个，则可以通过把区间细分并考察在各个分点上函数的变号情况，就能得到每个实根所属的小区间。若方程左边的函数 $f(x)$ 还满足定理的第二个条件时，根的隔离方法还可以简化。这时按照函数 $f(x)$ 的曲线通过其一阶导数 $f'(x)$ 的零点时 $f'(x)$ 通常要改变符号这一事实，只要判断在给定区间 $[a,b]$ 的端点以及 $f'(x)$ 的零点处 $f'(x)$ 的变号情况即可。

例 3-1 隔离方程 $x^4-4x-1=0$ 的根。

解 取 $f(x)=x^4-4x-1$，定义域为 $(-\infty,+\infty)$，则

$$f'(x)=4(x^3-1)=0$$

它只有一个零点 $x=1$，此时 $f(1)=1-4-1=-4<0$。

在 $(-\infty,1)$ 上，$f'(x)<0$，$f(x)$ 单调减小，而 $f(-1)=4$。所以 $f(x)$ 在 $(-\infty,-1]$ 上均大于零，即 $x\in(-\infty,-1]$ 时，$f(x)>0$。于是 $f(x)$ 在 $(-1,1)$ 上有一个根。

在 $(1,+\infty)$ 上，$f'(x)>0$，$f(x)$ 单调增加，而 $f(2)=7$，则在 $(2,+\infty)$ 上 $f(x)$ 恒大于零。于是 $f(x)$ 一定有一个零点在 $(1,2)$ 内。

这样就判定出原方程只有两个根，分别落入 $(-1,1)$ 与 $(1,2)$ 内。

3. 计算机法

这种方法的理论根据也是零点定理。若已知函数 $f(x)$ 在 $[a,b]$ 上连续且 $f(a)\cdot f(b)<0$，则为实现根的隔离，可根据实际情况设定步长 h，由 a 点出发逐步寻找各根所在的小区间。其具体做法是首先求取 $f(a)$ 和 $f(a+h)$，若 $f(a)\cdot f(a+h)<0$，则在小区间 $[a,a+h]$ 内必有一个根，否则无根；然后再令 $a=a+h$，并求新的 $f(a)$ 和 $f(a+h)$，检验 $f(a)\cdot f(a+h)<0$？这样一直搜索到 $b=a+h$ 为止，便可找到 $[a,b]$ 内的全部有根区间。在应用这种方法时，应合理选择步长 h，h 过大会将根漏掉；h 过小则计算步骤过多。

二、图解法求根的粗略近似值

由于方程 $f(x)=0$ 的实根就是函数 $f(x)$ 的图像与 x 轴的交点，这样就可以用描绘函

数 $y=f(x)$ 的图形的方法粗略地找到根的近似值。这种粗略的近似值可作为根进一步精确化的出发点。

有时候为了绘图方便，可把方程 $f(x)=0$ 转化为其等价方程形式：$f_1(x)=f_2(x)$，将 $f_1(x)$ 和 $f_2(x)$ 绘在同一个图上，其交点的横坐标即为 $f(x)=0$ 的根的粗略近似值。

例 3-2 用图解法求方程 $x+\ln x-2=0$ 的粗略近似根。

解 设 $f_1(x)=2-x$，$f_2(x)=\ln x$，其图形如图 3-1 所示，其交点的横坐标 $x=1.3$ 即为原方程的粗略近似根。

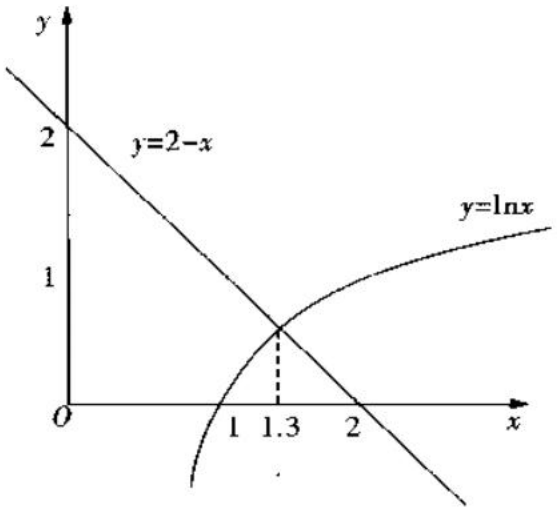


图 3-1

第二节 简单迭代法

一、简单迭代法的基本思想

简单迭代法亦称逐次逼近法。它是先将方程 $f(x)=0$ 写成等价形式

$$x=g(x) \tag{3-1}$$

然后将初值 x_0 代入式(3-1)右端求得 $x_1=g(x_0)$，再由 x_1 又求得 $x_2=g(x_1)$。如此继续下去，便可构造出近似解序列

$$x_k=g(x_{k-1}) \quad (k=1,2,\cdots,n)$$

直到满足条件 $|x_n-x_{n-1}|<\varepsilon$ 时，x_n 便为所求方程 $f(x)=0$ 的近似解。

这种求根的方法称为简单迭代法，其中，$x_k=g(x_{k-1})$ 称为迭代格式，$g(x)$ 称为迭代函数。

当 $\lim\limits_{k\to\infty}x_k=\lim\limits_{k\to\infty}(x_{k-1})$ 时，就称近似解序列收敛，否则称其发散。

显然，选择不同的迭代函数，将会产生不同的近似解序列 $\{x_k\}$，并不是所有的近似解序列都是收敛的。下面看一个例子。

例 3－3　已知方程 $x^3+4x^2-10=0$ 在[1,2]上有唯一的一个根，试用迭代法确定此根。

解　现用两种方法把所给的方程化成便于迭代的形式，以便于对迭代结果作比较。

(1) $x=g_1(x)=x^3+4x^2+x-10$；

(2) $x=g_2(x)=\left(\dfrac{10}{x+4}\right)^{\frac{1}{2}}$。

现将前 9 次迭代结果列表如下：

k	0	1	2	3	4
$x=g_1(x)$	1.5	3.875	112.123 04	$1.459\ 95\times10^6$	/
$x=g_2(x)$	1.5	1.348 399 73	1.367 376 37	1.364 957 01	1.365 264 75
k	5	6	7	8	9
$x=g_1(x)$	/	/	/	/	/
$x=g_2(x)$	1.365 225 59	1.365 230 58	1.365 229 94	1.365 230 02	1.365 230 01

从表中可以清楚地看出，用(1)式得到的解序列是发散的，不可能有极限，因而就得不到方程的根。用(2)式得到的数列是收敛的，相邻两项越来越接近，若用第 9 次得到的数 $x_9=1.365\ 230\ 01$ 作为根的近似值，则精度至少为 10^{-7}。

这样，究竟选取什么样的迭代函数才能保证所得的解序列 $\{x_k\}$ 收敛，对初始近似值有什么限制，以及收敛的速度、误差的估计等就成为应考虑的问题。为此，首先从几何上来观察迭代函数 $g(x)$ 应满足的条件。

二、迭代法的几何意义

迭代过程的收敛与发散情况可以从几何上得到比较明确的解释。现在直角坐标系上绘出 $y=x$ 与 $y=g(x)$ 的图形，如图 3－2 所示。其交点 M 的横坐标就是方程 $F(x)=0$ 的根 x^*。

如图 3－2(a)所示，假定从曲线 $y=g(x)$ 上某一个初始点 $S_0(x_0,g(x_0))$ 出发作平行于 x 轴的直线交 $y=x$ 于 T_1，再从 T_1 出发作平行于 y 轴的直线交 $y=g(x)$ 于 S_1，重复以上作法即得一条折线 $S_0T_1S_1T_2\cdots$，可以看出，折线前进的方向越来越接近于交点 M，而点 $S_0,S_1,S_2,\cdots$ 的横坐标 $x_0,x_1,x_2,\cdots$ 即是一个收敛于 x^* 的序列。[图 3－2(b)中的 $x_0,x_1,x_2,\cdots$ 是不收敛的情况]

$g(x)$ 的图形为什么具有这种性质呢？从图 3－2(a)中不难看出，曲线 $y=g(x)$ 在 x^* 的某邻域内变化率的绝对值 $|g'(x)|$ 不超过直线 $y=x$ 的斜率，即有 $|g'(x)|<1$。也只有此时，$\lim\limits_{k\to\infty}x_k=x^*$。

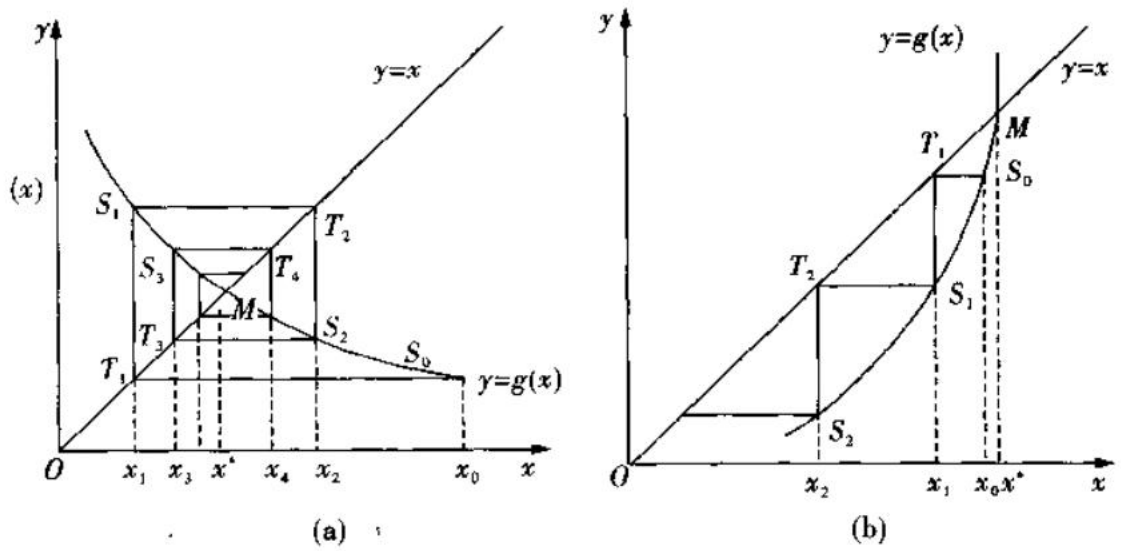


图 3－2

反之，若 $|g'(x)|>1$，由图 3－2(b) 可以看出数列 $\{x_k\}$ 是发散的。

由此可见，函数 $g(x)$ 在 x^* 邻域的变化率直接决定了数列 $\{x_k\}$ 的敛散性。

三、迭代法收敛性定理

以下定理给出了迭代法收敛的充分条件。这个定理可以用不同的方法证明，这里仅给出一种直接证明的方法。

定理 若函数 $g(x)$ 在 (a,b) 上连续可导，且当 $x\in[a,b]$ 时，$g(x)\in[a,b]$，同时存在一个小的正数 q，使得 $|g'(x)|\leqslant q<1$，则：

(1) $x=g(x)$ 在 $[a,b]$ 内有唯一解 x^*，且对于任意的 $x_0\in[a,b]$ 有 $\lim\limits_{k\to\infty}x_k=x^*$。

(2) $|x^*-x_k|\leqslant\dfrac{1}{1-q}|x_{k+1}-x_k|$，即 $|x^*-x_k|\leqslant\dfrac{q^k}{1-q}|x_1-x_0|$。

(3) $\lim\limits_{k\to\infty}\dfrac{x^*-x_{k+1}}{x^*-x_k}=g'(x^*)$。

证明 (1) 先证明根 x^* 的唯一性。

取 $f(x)=x-g(x)$，由于 $x\in[a,b]$ 时，$g(x)\in[a,b]$，则

$$f(a)=a-g(a)\leqslant 0,\quad f(b)=b-g(b)\geqslant 0$$

而

$$f'(x)=1-g'(x)>0$$

即 $f(x)$ 在 $[a,b]$ 上单调增加。

于是方程 $f(x)=0$ 在 $[a,b]$ 内有唯一解 x^*，且 $x^*=g(x^*)$。

再证明 $\lim\limits_{k\to\infty}x_k=x^*$。

由迭代公式知

$$x_k = g(x_{k-1}), x_{k+1} = g(x_k)$$

于是

$$x_{k+1} - x_k = g(x_k) - g(x_{k-1})$$

利用微分中值定理可得

$$(x_{k+1} - x_k) = (x_k - x_{k-1})g'(\xi), \quad \xi \in (x_{k-1}, x_k)$$

则

$$|x_{k+1} - x_k| \leqslant q|x_k - x_{k-1}|$$

其中，$|q| \geqslant |g'(\xi)|$。

令 $k = 1, 2, \cdots$，得到一组递推不等式：

$$|x_2 - x_1| \leqslant q|x_1 - x_0|$$

$$|x_3 - x_2| \leqslant q|x_2 - x_1| \leqslant q^2|x_1 - x_0|$$

$$\cdots\cdots\cdots\cdots$$

$$|x_k - x_{k-1}| \leqslant q^{k-1}|x_1 - x_0|$$

所以

$$|x_{k+1} - x_k| \leqslant q^k|x_1 - x_0|$$

现考察级数：

$$x_0 + (x_1 - x_0) + (x_2 - x_1) + \cdots + (x_k - x_{k-1}) + \cdots$$

显然，该级数前$(k+1)$项的和 $S_{k+1} = x_k$，且该级数的对应项小于或等于下面一个级数的对应项：

$$x_0 + (x_1 - x_0) + q(x_1 - x_0) + q^2(x_1 - x_0) + \cdots + q^{k-1}(x_1 - x_0) + \cdots$$

后一级数为公比 $q < 1$ 的几何级数，级数收敛，所以前一个级数也收敛，即

$$\lim_{k \to \infty} S_{k+1} = \lim_{k \to \infty} x_k = x^*$$

(2)由于

$$\begin{aligned} |x_{k+1} - x_k| &= |(x^* - x_k) - (x^* - x_{k+1})| \\ &\geqslant |x^* - x_k| - |x^* - x_{k+1}| \\ &\geqslant |x^* - x_k| - q|x^* - x_k| \\ &= (1 - q)|x^* - x_k| \end{aligned}$$

即

$$|x^* - x_k| \leqslant \frac{1}{1-q}|x_{k+1} - x_k| \leqslant \frac{q^k}{1-q}|x_1 - x_0|$$

(3)因为

$$x^* = g(x^*), \quad x_{k+1} = g(x_k)$$

所以

$$x^* - x_{k+1} = g(x^*) - g(x_k) = g'(\eta)(x^* - x_k)$$

其中，η 介于 x^* 和 x_k 之间。

于是

$$\frac{x^* - x_{k+1}}{x^* - x_k} = g'(\eta)$$

当 $k \to \infty$ 时，$x_k \to x^*$，$\eta \to x^*$。

故

$$\lim_{k \to \infty} \frac{x^* - x_{k+1}}{x^* - x_k} = g'(x^*)$$

利用 $|x^* - x_k| \leqslant \frac{q^k}{1-q}|x_1 - x_0|$ 可估计误差。当 q 较小时，即越靠近于零时，收敛性越好；当 q 靠近 1 时，收敛缓慢，而与初始近似值 x_0 的选择无关。另外，在收敛的前提下，只要 $|x_{k+1} - x_k|$ 充分小，x_{k+1} 就可以作为 x^* 的较好的近似值。因而实际中常常预先给定一个小正数 ε，以 $|x_{k+1} - x_k| \leqslant \varepsilon$ 来控制迭代过程。

例 3-4 用简单迭代法求解范德华方程，计算在 $T = 173$ K 和 $p = 50$ atm 下的氮气体积（要求精度 $\varepsilon = 10^{-6}$）。范德华方程为

$$\left(p + \frac{a}{V^2}\right)(V - b) = RT$$

式中，$a = 1.351\ (\text{atm} \cdot \text{L}^2 \cdot \text{mol}^{-2})$；$b = 38.64 \times 10^{-3}\ (\text{L} \cdot \text{mol}^{-1})$；$R = 82.06 \times 10^{-3}\ (\text{atm} \cdot \text{L} \cdot \text{mol}^{-1} \cdot \text{K}^{-1})$。

解 迭代函数可写为如下两种形式：

$$V = g_1(V) = RT / \left[p + \frac{a}{V^2}\right] + b$$

$$V = g_2(V) = \left[\frac{a(V-b)}{RT - p(V-b)}\right]^{\frac{1}{2}}$$

当取理想气体体积为

$$V_0 = \frac{RT}{p} = \frac{82.06 \times 10^{-3}}{50} \times 173 = 0.283\ 93\ (\text{L})$$

作为初值时，由于

$$|g_1'(V)| = \left|\frac{2aRT}{V^3\left(p + \frac{a}{V^2}\right)^2}\right| \leqslant |g_1'(0.283\ 93)| \leqslant 0.38 < 1$$

$$|g_2'(V)| = \left|\frac{1}{2}\left[\frac{a(V-b)}{RT - p(V-b)}\right]^{-\frac{1}{2}} \cdot \frac{aRT}{(RT - pV + pb)^3}\right| \leqslant |g_2'(0.283\ 93)| \geqslant 6.2 > 1$$

则应取 $V = g_1(V)$ 进行迭代，得 $V = 0.222\ 521\ 5$ (L)。

第三节　加速迭代收敛的 δ^2 —法

用简单迭代法产生的序列当 $|g'(x)|$ 接近于 1 时，收敛速度是缓慢的。实际中可用

不同的方法来改善迭代序列的收敛速度。现讨论常用的 δ^2—法。

一、δ^2—序列的定义及其性质

1. 定义

设有一序列 $\{x_k\}(k=0,1,2,\cdots)$，若二阶差分

$$\Delta^2 x_{k-1}=x_{k+1}-2x_k+x_{k-1}\neq 0$$

则把按公式

$$P(x_k)=\frac{\begin{vmatrix} x_{k-1} & x_k \\ x_k & x_{k+1} \end{vmatrix}}{\Delta^2 x_{k-1}}=\frac{x_{k-1}x_{k+1}-x_k^2}{x_{k+1}-2x_k+x_{k-1}}$$

构造的序列 $\{P(x_k)\}(k=1,2,\cdots)$ 叫做 δ^2—序列。

2. δ^2—序列 $\{P(x_k)\}$ 的性质

对于与 k 无关的元素 x，序列 $\{P(x_k)\}$ 的元素满足以下关系式：

$$P(x+x_k)=x+P(x_k)$$

事实上，

$$P(x+x_k)=\frac{\begin{vmatrix} x+x_{k-1} & x+x_k \\ x+x_k & x+x_{k+1} \end{vmatrix}}{\Delta^2(x+x_{k-1})}=\frac{x\Delta^2 x_{k-1}}{\Delta^2 x_{k-1}}+\frac{\begin{vmatrix} x_{k-1} & x_k \\ x_k & x_{k+1} \end{vmatrix}}{\Delta^2 x_{k-1}}$$

即

$$P(x+x_k)=x+P(x_k)$$

按此性质可得如下定理。

定理 若序列 $\{x_k\}$ 按几何级数收敛于 x^*，即 $x_k-x^*=Cq^k$，$0<q<1$，C 为常数，那么序列 $\{p(x_k)\}$ 的所有项都等于 x^*，即 $P(x_k)=x^*$。

证明 由于

$$P(Cq^k)=P(-x^*+x_k)=-x^*+P(x_k)$$

但

$$P(Cq^k)=\frac{\begin{vmatrix} Cq^{k-1} & Cq^k \\ Cq^k & Cq^{k+1} \end{vmatrix}}{\Delta^2(Cq^{k-1})}=0$$

所以

$$x^*=P(x_k)\quad(k=1,2,\cdots)$$

这就是用 δ^2—法加速的根据。由于用简单迭代法产生的序列当满足收敛性定理的条件时，通常是一个接近于按几何级数收敛的序列，所以用 δ^2—法可加速其收敛速度。

二、用 δ^2—法加速斯蒂芬算法

斯蒂芬算法的基本思想是：进行两步简单迭代，作一次加速，产生一个迭代与加速的混合序列。具体地说，就是从一个初始点 x_{k-1} 出发，用简单迭代公式 $x_k=g(x_{k-1})$ 算出 x_k，

x_{k+1}，再利用已得到的 3 个点 x_{k-1}, x_k, x_{k+1}，按 δ^2—法加速公式算出第 4 个点 x_{k+2}，当 $|x_{k+2}-x_{k+1}|\leqslant\varepsilon$ 时，终止计算，否则，令 $x_{k-1}=x_{k+2}$，再重复以上过程：

为了便于计算，现导出计算 x_{k+2} 的表达式：

$$x_{k+2}=P(x_k)=\frac{x_{k-1}x_{k+1}-x_k^2}{\Delta^2 x_{k-1}}$$

$$=\frac{x_{k-1}x_{k+1}-x_k^2-x_{k+1}^2+2x_{k-1}x_k+x_{k+1}^2-2x_{k+1}x_k}{\Delta^2 x_{k-1}}$$

$$=\frac{x_{k+1}(x_{k+1}-2x_k+x_{k-1})}{\Delta^2 x_{k-1}}-\frac{(x_{k+1}-x_k)^2}{\Delta^2 x_{k-1}}$$

即

$$x_{k+2}=x_{k+1}-\frac{(x_{k+1}-x_k)^2}{\Delta^2 x_{k-1}}$$

或

$$x_{k+2}=P(x_k)=\frac{x_{k-1}x_{k+1}-x_k^2-x_{k-1}^2+2x_{k-1}x_k+x_{k-1}^2-2x_{k-1}x_k}{\Delta^2 x_{k-1}}$$

$$=\frac{x_{k-1}(x_{k+1}-2x_k+x_{k-1})}{\Delta^2 x_{k-1}}-\frac{(x_k-x_{k-1})^2}{\Delta^2 x_{k-1}}$$

即

$$x_{k+2}=x_{k+1}-\frac{(x_k-x_{k-1})^2}{\Delta^2 x_{k-1}}$$

综上所述，斯蒂芬算法的步骤为：

(1)给定迭代的初始值 x_0 和要求的精度 ε。

(2)由 $x_k=g(x_{k-1})$ 计算出 x_1, x_2。

(3)利用 $x_3=x_2-\frac{(x_2-x_1)^2}{\Delta^2 x_0}$ 或 $x_3=x_0-\frac{(x_1-x_0)^2}{\Delta^2 x_0}$ 计算出 x_3。

(4)若 $|x_3-x_2|\leqslant\varepsilon$，输出 $x=x_3$，终止计算；否则，令 $x_0=x_3$，回第(2)步。

第四节　韦格斯坦加速迭代法

简单迭代法受收敛条件 $|g'(x)|<1$ 的限制，且收敛速度较慢，韦格斯坦法则是对简单迭代法的一种改进，它不仅不受 $|g'(x)|<1$ 的限制，并且加快了收敛速度。在化工流程模拟中，这种方法得到了广泛应用。

为了直观地从几何图形分析来建立迭代格式，现仍将方程 $f(x)=0$ 改写为 $x=g(x)$。

如图 3-3 所示，从任意的两个初始点 x_0, x_1 出发作 x 轴的垂线，交曲线 $y=g(x)$ 于 $(x_0, g(x_0))$ 和 $(x_1, g(x_1))$ 两点。过 $(x_0, g(x_0))$，$(x_1, g(x_1))$ 两点连一直线，该直线的斜率为

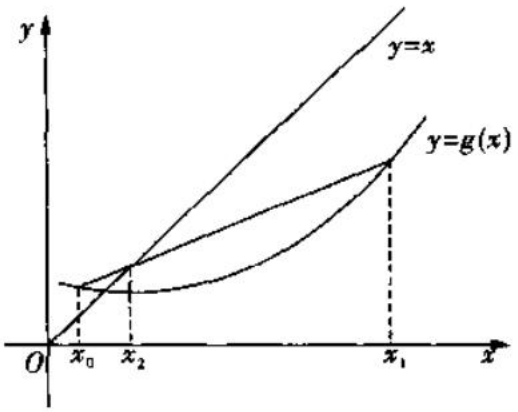


图 3-3

$$S=\frac{g(x_1)-g(x_0)}{x_1-x_0}$$

它与直线 $y=x$ 相交于 (x_2,x_2) 点。此交点的横坐标 x_2 即为韦格斯坦法迭代的一步解。然后由 x_1,x_2 出发重复上述步骤，直至找到满足精度要求的解。

依照上述步骤，可导出韦格斯坦法的迭代格式。对于任给的两个初始点 $(x_{k-1},g(x_{k-1}))$ 和 $(x_k,g(x_k))$，连接两点的直线方程为

$$y-g(x_{k-1})=\frac{g(x_k)-g(x_k-1)}{x_k-x_{k-1}}(x-x_{k-1})$$

它与直线 $y=x$ 相交于点 (x_{k+1},x_{k+1})，则

$$x_{k+1}-g(x_{k-1})=\frac{g(x_k)-g(x_{k-1})}{x_k-x_{k-1}}(x_{k+1}-x_{k-1})$$

整理得

$$x_{k+1}=\frac{x_kg(x_{k-1})-x_{k-1}g(x_k)}{(x_k-x_{k-1})-[g(x_k)-g(x_{k-1})]}\quad(k=1,2,\cdots)$$

由图 3-3 可以看出，若 x_0,x_1 两点选择不当，致使过 $(x_0,g(x_0))$ 和 $(x_1,g(x_1))$ 两点的直线的斜率 $S=1$，则它与直线 $y=x$ 平行，那就找不到交点 x_2 了。因此，应注意选择两个初始点，避免出现这种情况。

为了与简单迭代法相比较，现考察韦格斯坦法的迭代格式：

$$x_{k+1}=g(x_k)+S(x_{k+1}-x_k)$$

其中，$S=[g(x_k)-g(x_{k-1})]/(x_k-x_{k-1})$。

于是

$$x_{k+1}=\frac{1}{1-S}g(x_k)-\frac{S}{1-S}x_k$$

令 $C=\frac{1}{1-S}$，则 $1-C=-\frac{S}{1-S}$。

故

$$x_{k+1}=(1-C)x_k+Cg(x_k)$$

当 $C=1$ 时，即为简单迭代法。当 $0<C<1$ 时，则变为有阻尼的顺序迭代法。通常当 $C<1$ 时能稳定收敛，但较慢。当 $C>1$ 时能加速收敛，但易导致不稳定。为了既加速收敛又避免不稳定，常规定 $1<C<6$，这时的韦格斯坦法称为加界的韦格斯坦法。

由韦格斯坦迭代法的迭代公式可知，它不仅需要两个初始值，而且以后的每次计算都要用到前两次的计算结果。这样韦格斯坦加速迭代法的计算步骤可概括为：

(1)给出两个初值 x_0，x_1 和精度要求 ε。

(2)由 $x_2=\dfrac{x_1g(x_0)-x_0g(x_1)}{(x_1-x_0)-[g(x_1)-g(x_0)]}$ 求得 x_2。

(3)若 $|x_2-g(x_2)|\leqslant\varepsilon$，则输出 $x=x_2$，终止计算；否则，令 $x_0=x_1$，$x_1=x_2$，回第(2)步。

第五节　牛顿法

牛顿法又称切线法，本质上仍属迭代法，它对于便于用解析法求导数的函数方程求根是一种有效的方法，特别适用于高次代数方程和超越函数方程，其特点是程序简单，只要初值适当，收敛速度快。

一、方法概述

当函数 $f(x)$ 在 $[a,b]$ 上具有连续的二阶导数时，可利用曲线 $f(x)$ 上某点处的切线来代替曲线求方程 $f(x)=0$ 的根。

如图 3-4 所示，过曲线 $y=f(x)$ 上的一点 $(x_k,f(x_k))$ 作曲线的切线，切线方程为

$$y-f(x_k)=f'(x_k)(x-x_k)$$

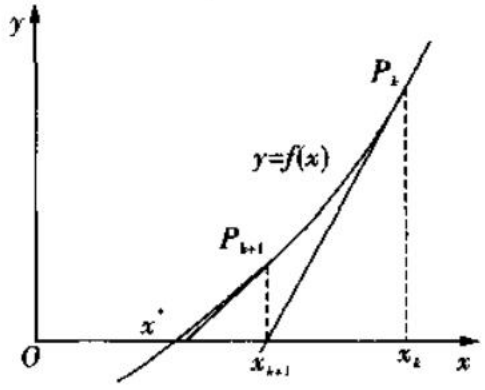


图 3-4

该切线与 x 轴的交点即为 $(x_{k+1},0)$，于是得牛顿法的迭代格式：

$$x_{k+1}=x_k-\frac{f(x_k)}{f'(x_k)}$$

可见牛顿法就是函数 $g(x)=x-\dfrac{f(x)}{f'(x)}$ 时的简单迭代法。这种 $g(x)$ 的特定形式决定了它有较高的收敛速度。

二、牛顿法收敛性定理

这里不加证明直接给出其收敛的判别法以及相应的误差估计式。

定理 若方程 $f(x)=0$ 中的 $f(x)$ 二阶连续可导，且对于根的初始近似值 x_0 满足：

(1) $f'(x_0)\neq 0$，$\dfrac{1}{|f'(x_0)|}\leqslant B$；

(2) $\left|\dfrac{f(x_0)}{f'(x_0)}\right|\leqslant D$；

(3) 在 x_0 的领域 $|x-x_0|\leqslant 2D$ 内，$|f'(x)|\leqslant K$。

则当 $h=BD\cdot K\leqslant\dfrac{1}{2}$ 时，由迭代格式

$$x_{k+1}=x_k-\frac{f(x_k)}{f'(x_k)}$$

得到的迭代序列 $\{x_k\}$ 收敛于 x^*，且 $f(x^*)=0$，其收敛速度为

$$|x^*-x_k|\leqslant\frac{1}{2^{n-1}}(2h)^{2^k-1}D$$

由此定理可见，牛顿法的收敛速度是超几何速度。因为它是以 2^k 作为指数，而不是以 k 作为指数。

在实际使用牛顿法时，为了简化计算，常用下述方法判别收敛性以及选取初值 x_0。

收敛判别法 若 $f(x)$ 在 (a,b) 上有连续的二阶导数，且满足条件：

(1) $f(a)\cdot f(b)<0$；

(2) $f'(x)\neq 0$，$f''(x)$ 在 (a,b) 上不变号；

(3) 初值 $x_0\in[a,b]$，且 $f(x_0)f''(x_0)>0$。

则由 $x_{k+1}=x_k-\dfrac{f(x_k)}{f'(x_k)}$ 给出的迭代序列 $\{x_k\}$ 收敛于方程 $f(x)=0$ 在 $[a,b]$ 上的唯一解。

在使用牛顿法时，一定要注意其对初始值 x_0 的要求，即保证 $f(x_0)f''(x_0)>0$，否则，将导致计算失败。

事实上，设 x_0 是 $f(x)=0$ 的一初始近似值，且在 (x_0,x) 内方程有唯一根，同时 $f''(x)$ 保持不变号。由泰勒公式得

$$f(x)=f(x_0)+(x-x_0)f'(x_0)+\frac{1}{2}(x-x_0)^2f''(\xi)\quad(x_0<\xi<x)$$

由于 $f(x)=0$，便有

$$x=x_0-\frac{f(x_0)}{f'(x_0)}-\frac{1}{2}(x-x_0)^2\frac{f''(\xi)}{f'(x)}$$

又由牛顿法的迭代格式可知

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}, \quad x_1 - x_0 = -\frac{f(x_0)}{f'(x_0)}$$

于是

$$x - x_1 = -\frac{1}{2}(x - x_0)^2 \frac{f''(\xi)}{f'(x_0)}$$

则

$$\frac{x - x_1}{x_1 - x_0} = \frac{1}{2}(x - x_0)^2 \frac{f''(\xi)}{f(x_0)}$$

在$[x_0, x]$内，当 $\operatorname{sgn} f''(x) = \operatorname{sgn} f(x_0)$时，有 $\operatorname{sgn}(x - x_1) = \operatorname{sgn}(x_1 - x_0)$，于是 x_1 位于 x 与 x_0 之间。

牛顿法过程简单，使用方便，其计算步骤为：

(1)给定初值 x_0 和计算精度 ε。

(2)计算 $f(x_0)$，$f'(x_0)$。

(3)由 $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$求得 x_1。

(4)若 $f(x_1) = 0$ 或 $|x_1 - x_0| \leqslant \varepsilon$，输出 $x = x_1$，停止计算；否则，令 $x_0 = x_1$，回第(2)步。

例 3-5 利用牛顿法求方程 $x^3 - x^2 - 4x - 7 = 0$ 在$[3,4]$内的实根，精度要求 $\varepsilon = 10^{-5}$。

解 取 $f(x) = x^3 - 2x^2 - 4x - 7$，则

$$f(3) = -10, \quad f(4) = 9, \quad f'(x) = 3x^2 - 4x - 4$$

在$[3,4]$内，$f'(x) > 0$，方程仅有一个根，此时

$$f''(x) = 6x - 4 > 0$$

则应取 $x_0 = 4$ 作为初值，迭代 4 次得 $x = 3.631\ 98$。

例 3-6 通过在温度范围为 200 ~ 600 K 之间的精确测量，给出甲苯胺蒸气压(p, mmHg)与温度(T, K)关系的经验公式为：

$$\ln p = 54.869\ 7 - \frac{8\ 013.7}{T} - 5.081 \ln T$$

试求 $p = 760$ mmHg 下的沸点，计算精度 $\varepsilon = 10^{-4}$。

解 对 $T \in [200, 600]$，$p = 760$ mmHg，取

$$f(T) = 54.869\ 7 - \frac{8\ 013.7}{T} - 5.081 \ln T - \ln p$$

则

$$f(T) = 48.236\ 4 - \frac{8\ 013.7}{T} - 5.081 \ln T$$

$$f(200) = -18.752\ 9, \quad f(600) = 2.327\ 4$$

$$f'(T)=\frac{8\,013.7}{T^2}-\frac{5.081}{T}>0$$

$$f''(T)=\frac{5.081}{T^2}-\frac{16\,027.4}{T^3}<0$$

取 $T_0=400$ K，迭代 5 次得 $T=476.009\,9$ K。

第六节　弦位法

牛顿法虽然收敛速度快，但需求出 $f'(x)$，当函数 $f(x)$ 的 $f'(x)$ 不易求得时，则可采用弦位法解方程 $f(x)=0$。弦位法的基本思想与牛顿法相似，即用直线代替曲线 $f(x)$，其差别在于弦位法是采用曲线上两点间的割线，而不是某点的切线。

若用过点 $(x_0,f(x_0))$ 和 $(x_1,f(x_1))$ 间的割线代替曲线 $y=f(x)$（见图 3-5），即用

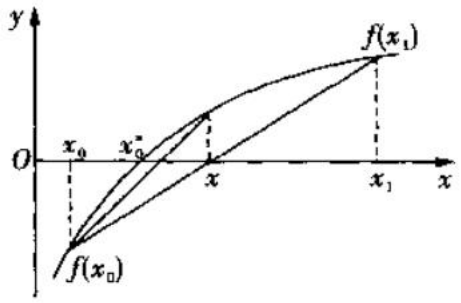


图 3-5

$$f(x_1)+\frac{f(x_1)-f(x_0)}{x_1-x_0}(x-x_1)=0$$

代替原方程，解得

$$x=x_1-\frac{x_1-x_0}{f(x_1)-f(x_0)}f(x_1)$$

对任给两个初值 x_{k-1},x_k，迭代格式为

$$x_{k+1}=x_k-\frac{x_k-x_{k-1}}{f(x_k)-f(x_{k-1})}f(x_k)\quad(k=1,2,\cdots)$$

当由初值 $x_0,x_1\in[a,b]$ 出发，按弦位法迭代公式求得近似根 x_2 之后，其函数值 $f(x_2)$ 可有 3 种情况：

第一种情况是 $f(x_2)=0$，则 x 便为方程 $f(x)=0$ 的精确解。

第二种情况是 $f(x_2)\cdot f(x_1)>0$，这表明根 $x^*\in[x_0,x_2]$，以 x_2 代替 x_1，$f(x_2)$ 代替 $f(x_1)$ 继续迭代。

第三种情况是 $f(x_2)\cdot f(x_1)<0$，这表明根 $x^*\in[x_2,x_1]$，则以 x_2 和 $f(x_2)$ 代替 x_0 和 $f(x_0)$ 继续迭代。

可以证明，当 $f(x)$，$f'(x)$ 和 $f''(x)$ 在解 x^* 的某个充分小的邻域 $R=\{x\,|\,|x-x^*|\leqslant\delta|$

内连续，且$f'(x^*)\neq 0$时，对任选初值$x_0, x_1 \in R$，由弦位法产生的解序列$\{x_k\}(k=1,2,\cdots)$收敛于x^*。

弦位法虽比牛顿法收敛速度稍慢，但在每次迭代中只需计算一次函数值，不必求函数的导数，且对初值x_0和x_1要求不甚苛刻，是工程计算中常用的有效算法之一。

由前所述可知，弦位法的计算步骤为：

(1)给定x_0, x_1和ε, δ。

(2)计算$f(x_0), f(x_1)$。

(3)若$f(x_0)f(x_1)>0$，则$x_0 = x_0 - \delta, x_1 = x_1 + \delta$，回第(2)步。

(4)由$x_2 = x_1 - \dfrac{x_1 - x_0}{f(x_1) - f(x_0)} f(x_1)$，求$x_2$。

(5)若$|f(x_2)| \leqslant \varepsilon$，输出$x = x_2$，终止计算。

(6)如果$f(x_2) \cdot f(x_1) < 0$，则令$x_0 = x_1, f(x_0) = f(x_1), x_1 = x_2, f(x_1) = f(x_2)$，回第(4)步。

(7)如果$f(x_2) \cdot f(x_1) > 0$，则令$x_1 = x_2, f(x_1) = f(x_2)$，回第(4)步。

例3-7 利用弦位法求方程$x^3 + 1.1x^2 + 0.9x - 1.4 = 0$在[0,1]内的一个根，计算精度$\varepsilon = 10^{-4}$。

解 取

$$f(x) = x^3 + 1.1x^2 + 0.9x - 1.4$$

则

$$f(0) = -1.4, \quad f(1) = 1.6$$

$$x_2 = 1 - \frac{1}{1.6 - (-1.4)} \times 1.6 = 0.2$$

$$f(0.2) = -1.168$$

新的计算区间为[0.2,1]，求得$x_3 = 0.5376$。同样方法经8次迭代求得$x = 0.6706$。

第七节 二分法

若$f(x)$在$[a,b]$上单调连续，且$f(a) \cdot f(b) < 0$，则可以用把区间$[a,b]$逐渐分半的方法来求方程$f(x) = 0$的根，这种方法称为二分法。

现结合图3-6来说明二分法的基本思想。先算出区间$[a,b]$的中点坐标$c = \dfrac{a+b}{2}$，判断c是否为方程$f(x) = 0$的满足一定精度的根。若是，则结束二分过程；否则，根一定落在半区间$[a,c]$或$[c,b]$内。记新的含根区间为$[a_1,b_1]$，显然这个区间满足二分法的条件，即$f(a_1) \cdot f(b_1) < 0$，把新的区间$[a_1,b_1]$二等分得$c_1 = (a_1 + b_1)/2$。重复以上过程，便得到一个区间套序列：

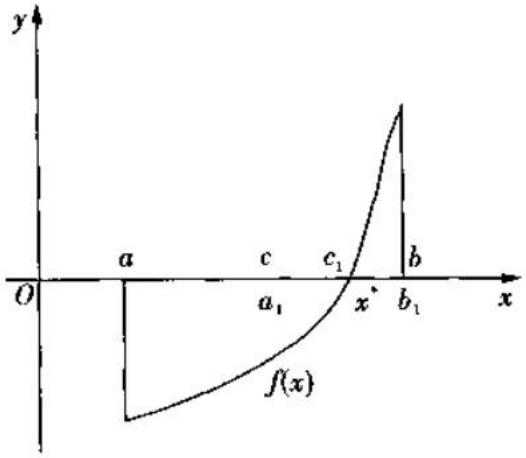


图 3-6

$$[a,b]\supset[a_1,b_1]\supset[a_2,b_2]\supset\cdots\supset[a_k,b_k]\supset\cdots$$

依此分法显然有

$$f(a_k)\cdot f(b_k)<0 \quad (k=1,2,\cdots)$$

并且

$$b_k-a_k=\frac{1}{2^k}(b-a)$$

由于 $a_1,a_2,\cdots$ 为非减序列，$b_1,b_2,\cdots$ 为非增序列。由上式可知它们有共同的极限

$$x^*=\lim_{k\to\infty}a_k=\lim_{k\to\infty}b_k$$

由于 $f(x)$ 在 $[a,b]$ 上连续，所以

$$\lim_{k\to\infty}f(a_k)\cdot f(b_k)=0$$

即

$$f(x^*)=0$$

这说明序列的极限 x^* 是方程的根。

由前述容易看出根的近似值有如下估计式：

$$|x^*-x_k|\leqslant\frac{1}{2^k}(b-a)$$

依照以上讨论可概括出以下定理。

定理 若 $f(x)=0$ 中 $f(x)$ 在 $[a,b]$ 上单调连续，且 $f(a)\cdot f(b)<0$，则由二分法产生的序列 $\{x_k\}$ 收敛于方程的根且有误差估计式

$$|x_k-x^*|\leqslant\frac{1}{2^k}(b-a) \quad (k=1,2,\cdots)$$

二分法的优点是程序简单，且只要函数 $f(x)$ 连续即可，有了区间的两端点就能计算，不用选取初始近似值。但对于偶重根的情况，二分法是不能使用的。这是因为在根的邻域内不存 $f(a)\cdot f(b)<0$ 这个条件。二分法的计算步骤可概括为：

(1)先确定出$f(x)$连续的有根区间$[a,b]$,使满足$f(a)\cdot f(b)<0$,并给出精度ε。

(2)令$c=\dfrac{a+b}{2}$,计算$f(c)$。

(3)若$|f(c)|\leqslant\varepsilon$,则输出$x=c$,终止计算。

(4)若$f(a)\cdot f(c)>0$,则令$a=c$,$f(a)=f(c)$,回第(2)步;若$f(a)\cdot f(c)<0$,则令$b=c$,$f(b)=f(c)$,回第(2)步。

第八节　迭代法的收敛阶

为了表明迭代序列$\{x_k\}$的收敛速度,引入迭代序列收敛阶的概念,它是衡量某迭代法优劣的标志之一。

定义　设$\{x_k\}$为收敛于x^*的序列,如果误差$e_k=x_k-x^*$满足

$$\lim_{k\to\infty}\frac{|e_{k+1}|}{|e_k|^p}=\lim_{k\to\infty}\frac{|x_{k+1}-x^*|}{|x_k-x^*|^p}=m\neq 0$$

其中,$p\geqslant 1$,则称序列$\{x_k\}$为p阶收敛;当$p=1$,且$0<m<1$时,序列$\{x_k\}$称为线性收敛;当$p>1$时,称序列$\{x_k\}$为超线性收敛;当$p=2$时,称为二阶收敛或平方收敛。

若由迭代格式产生的近似解序列$\{x_k\}$是p阶收敛,则称该迭代过程为p阶收敛。

显然,p的大小反应了迭代法收敛的快慢,p越大,收敛速度就越快。因此,收敛阶p是对迭代法收敛速度的定量表示。

例如,对于线性收敛,$p=1$,$0<m<1$,则

$$\lim_{k\to\infty}\left|\frac{e_{k+1}}{e_k}\right|=\lim_{k\to\infty}\frac{|x_{k+1}-x^*|}{|x_k-x^*|}=m<1$$

于是有

$$|e_{k+1}|\leqslant m|e_k|\leqslant m^2|e_{k-1}|\leqslant\cdots\leqslant m^{k+1}|e_0|$$

迭代次数为

$$k\leqslant[\lg|e_{k+1}|-\lg|e_0|)/\lg m]-1$$

对于二阶收敛,$p=2$,则$\lim\limits_{k\to\infty}\dfrac{|e_{k+1}|}{|e_k|^2}=m$。

于是

$$|e_{k+1}|\leqslant m|e_k|^2\leqslant m^3|e_{k-1}|^4\leqslant\cdots\leqslant(m|e_0|)^{2(k+1)-1}|e_0|$$

迭次次数k可由下式计算:

$$(2^{k+1}-1)\lg(m|e_0|)\geqslant\lg|e_{k+1}|-\lg|e_0|$$

这样,在相同的初值和精度要求下,二阶收敛迭代过程要比线性收敛迭代过程快得多。例如,取$|e_0|=|x_1-x_0|=1$,$m=0.75$,要求解的误差限$|e_{k+1}|\leqslant\varepsilon=10^{-8}$。线性收敛迭代过程需要迭代63次,而二阶收敛迭代过程只需迭代5次。

可以证明,简单迭代法是线性收敛,弦位法是1.618阶收敛,牛顿法和δ^2—法加速迭

代为二阶收敛。下面仅对简单迭代法及牛顿法作出证明。

对于简单迭代法，设 $m=\max\limits_{a\leqslant x\leqslant b}|g'(x)|<1$，而

$$x^*=g(x^*),\quad x_{k+1}=g(x_k)\quad (k=0,1,2,\cdots)$$

两式相减，并应用微分中值定理有

$$|x^*-x_{k+1}|=|g(x^*)-g(x_k)|=|g'(\xi)||x^*-x_k|\leqslant m|x^*-x_k|$$

其中，$\xi\in(a,b)$。

则

$$\lim_{k\to\infty}\frac{|x^*-x_{k+1}|}{|x^*-x_k|}=\lim_{k\to\infty}\frac{|e_{k+1}|}{|e_k|}\leqslant m<1$$

即 $p=1$，说明简单迭代法是线性收敛。

对于牛顿法，由于 $f'(x^*)\neq 0$。现对函数 $f(x)$ 在 x_k 处作泰勒展开：

$$f(x)=f(x_k)+f'(x_k)(x-x_k)+\frac{1}{2}f''(\xi)(x-x_k)^2$$

其中，ξ 介于 x 与 x_k 之间，当 $x=x^*$ 时，$f(x^*)=0$。

则

$$f(x_k)+f'(x_k)(x^*-x_k)+\frac{1}{2}f''(\xi)(x^*-x_k)^2=0$$

于是

$$x^*=\left[x_k-\frac{f(x_k)}{f'(x_k)}\right]-\frac{1}{2}\frac{f''(\xi)}{f'(x_k)}(x^*-x_k)^2$$

因此

$$x^*-x_{k+1}=-\frac{1}{2}\frac{f''(\xi)}{f'(x_k)}(x^*-x_k)^2$$

$$|x^*-x_{k+1}|=\frac{1}{2}\left|\frac{f''(\xi)}{f'(x_k)}\right||x^*-x_k|^2$$

故

$$\lim_{k\to\infty}\frac{|x^*-x_{k+1}|}{|x^*-x_k|^2}=\lim_{k\to\infty}\frac{|e_{k+1}|}{|e_k|^2}=\frac{1}{2}\left|\frac{f''(\xi)}{f'(x_k)}\right|\quad (\text{当 } k\to\infty \text{ 时，} x_k\to x^*,\xi\to x^*)$$

所以牛顿法在 x^* 附近为二阶收敛。当初值满足 $f(x_0)f''(x_0)>0$，且较接近 x^* 时，迭代过程很快。

习　题

1. 试在[0,2]之间确定方程 $x^4-3x+1=0$ 的两个实根所在区间。

2. 用图解法确定方程 $\sin x-\frac{x}{2}=0$ 的唯一正根的粗略近似值。

3. 对方程 $3x^2 - e^x = 0$，确定正根所在的区间 $[a,b]$ 和迭代函数 $g(x)$。

4. 求满足流动方程 $8\,820D^5 - 2.31D - 0.646\,5 = 0$ 的管径 D。（要求精度 $\varepsilon = 10^{-3}$）

5. 用亚硫酸钠吸收空气中的二氧化硫时，得出使每年花费最小时的塔径满足以下关系式：

$$24D^{0.64} - \frac{0.912\,8}{D^5} + 32\,554.46D + 3\,224.46D^{0.6} = 0$$

试用牛顿法求此时的塔径 D。（要求精度 $\varepsilon = 10^{-4}$）

6. 试用斯蒂芬算法求方程 $x^3 + 4x^2 - 10 = 0$ 的根。

7. 设 $P > 0$，写出牛顿法求 $\sqrt{P}$ 的迭代公式，并试算 $\sqrt{7}$ 的近似值。（要求精度 $\varepsilon = 10^{-6}$）

8. 试用韦格斯坦法和弦位法求方程 $x = 2^{-x}$ 在 $\left[\frac{1}{3},1\right]$ 上的唯一解。（要求精度 $\varepsilon = 10^{-5}$）

9. 试用迭代法、牛顿法、弦位法、二分法确定方程 $\cos x - x = 0$ 的近似根，要求精度 $\varepsilon = 10^{-6}$，并比较迭代次数。

10. 试用弦位法求 $x^x = 10^4$ 的根，要求计算精度 $\varepsilon = 10^{-4}$。

第四章　线性代数方程组的解法

许多工程技术问题，最终均归结到对一个线性方程组的求解。在化学工程中求解线性方程组的场合很多，如多组分体系的物料衡算、计算各种化合物的物理化学性质、分离装置的平衡级模拟以及稳态动力学计算等都需要解线性方程组。

至于线性代数方程组的解法，原则上讲，用代数学中熟知的克莱姆法则即可解决，然而实践上这种方法的计算量太大，解一个由 n 个方程组成的线性方程组要计算 $(n+1)$ 个行列式，而每个行列式的计算需要进行 $(n-1)n!$ 次乘法运算，最后还需要进行 n 次除法运算，所以总共乘除法的总次数为 $(n^2-1)n!+n$。

由此可见，当 n 增大时，计算次数会迅速增加，以至于使计算成为不可能。正因为如此，许多更为有效的数值方法才应运而生。下面就一些常用的线性代数方程组的数值解法逐一进行讨论。

第一节　高斯消去法

一、方法的基本思想

对于线性方程组

$$\left.\begin{aligned} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n&=b_1\\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n&=b_2\\ &\cdots\cdots\\ a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n&=b_n \end{aligned}\right\} \tag{4-1}$$

其矩阵形式为

$$AX=B$$

其中，$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$ 为系数矩阵；$X=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 为未知向量；$B=(b_1,b_2,\cdots,b_n)^{\mathrm{T}}$ 为右端向量，又称为右端自由项。

所谓高斯消去法就是通过一系列的初等变换，把系数矩阵 A 化为一个三角矩阵（通常化为上三角矩阵）F，同时右端向量变为 G，即方程组化为

$$FX=G \tag{4-2}$$

由于 F 为上三角矩阵，其最后一个方程也就只含一个未知数 x_n，向前每个方程较其后一个方程多一个未知数。这样就可以从最后一个方程开始，用逐步回代的方法求得方程组(4-1)的解。整个方法分为以下两个过程。

(1)正过程：通过初等变换，进行消元，使矩阵 A 化为三角矩阵 F，自由项 B 化为 G，得方程组 $FX=G$。

(2)逆过程：从 $FX=G$ 的最后一个方程逐步回代得方程组的解。

为了便于在计算机上实现求解过程，下面给予其较详细的说明。

在方程组 $AX=B$ 中，若 $a_{11}\neq 0$，则以 a_{11} 除第一个方程两边各项得

$$x_1+\frac{a_{12}}{a_{11}}x_2+\frac{a_{13}}{a_{11}}x_3+\cdots+\frac{a_{1n}}{a_{11}}x_n=\frac{b_1}{a_{11}}$$

若将 b_i 记为 $a_{i\,n+1}(i=1,2,\cdots,n)$；即 a_{1j}/a_{11} 仍记为 a_{1j}，即

$$a_{1j}=a_{1j}/a_{11}\quad (j=1,2,\cdots,n+1)$$

则得

$$x_1+a_{12}x_2+a_{13}x_3+\cdots+a_{1n}x_n=a_{1n+1}$$

再利用上式把方程组 $AX=B$ 中其他方程中含 x_1 的项均化为零。方法是将上式两边同乘以 $a_{i1}(i=2,3,\cdots,n)$，并与第 i 个方程相减，便把第 i 个方程中的 x_1 项消去，所得新方程为

$$(a_{i2}-a_{i1}a_{12})x_2+(a_{i3}-a_{i1}a_{13})x_3+\cdots+(a_{in}-a_{i1}a_{1n})x_n=(a_{i\,n+1}-a_{i1}a_{1n+1})$$

若将 $a_{ij}-a_{i1}a_{1j}$ 仍记为 a_{ij}，即

$$a_{ij}=a_{ij}-a_{i1}a_{1j}\quad (i=2,3,\cdots,n;j=2,3,\cdots,n+1)$$

可得

$$a_{i2}x_2+a_{i3}x_3+\cdots+a_{in}x_n=a_{i\,n+1}\quad (i=2,3,\cdots,n)$$

从而，除第一个方程外，得到了一个 $(n-1)$ 阶的方程组

$$\begin{cases} a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = a_{2\,n+1} \\ a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n = a_{3\,n+1} \\ \cdots\cdots \\ a_{n2}x_2 + a_{n3}x_3 + \cdots + a_{nn}x_n = a_{n\,n+1} \end{cases}$$

以上过程称为第一次消元。从第一次消元得到的降阶方程组出发，反复施行同样的方法，经第 k 次消元后可以得方程

$$x_k + a_{k\,k+1}x_{k+1} + a_{k\,k+2}x_{k+2} + \cdots + a_{kn}x_n = a_{k\,n+1}$$

及一个 $(n-k)$ 阶的方程组

$$\begin{cases} a_{k+1\,k+1}x_{k+1} + a_{k+1\,k+2}x_{k+2} + \cdots + a_{k+1\,n}x_n = a_{k+1\,n+1} \\ a_{k+2\,k+1}x_{k+1} + a_{k+2\,k+2}x_{k+2} + \cdots + a_{k+2\,n}x_n = a_{k+2\,n+1} \\ \cdots\cdots \\ a_{n\,k+1}x_{k+1} + a_{n\,k+2}x_{k+2} + \cdots\cdots + a_{nn}x_n = a_{n\,n+1} \end{cases}$$

其中，$a_{kj} = \dfrac{a_{kj}}{a_{kk}}$ $(j = k+1, k+2, \cdots, n+1)$；$a_{ij} = a_{ij} - a_{ik}\dfrac{a_{kj}}{a_{kk}}$ $(i = k+1, k+2, \cdots, n+1)$。

这两式为消元过程的一般计算公式。只要令 $k = 1, 2, 3, \cdots, n-1$，即可实现全部消元任务，把原方程组 $AX = B$ 化为等价的上三角形式的方程组

$$\begin{bmatrix} 1 & a_{12} & \cdots & \cdots & a_{1n} \\ 0 & 1 & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{1\,n+1} \\ a_{2\,n+1} \\ \vdots \\ a_{n\,n+1} \end{bmatrix}$$

经上述步骤，方程组已经达到了消元的目的，随后即可进行回代，求得解向量

$$\begin{cases} x_n = a_{n\,n+1}/a_{nn} \\ x_i = a_{i\,n+1} - \sum\limits_{j=i+1}^{n} a_{ij}x_j \quad (i = n-1, n-2, \cdots, 1) \end{cases}$$

高斯消去法的优点是计算步骤比较单一，便于在计算机上实现。但其亦存在严重的不足，它不仅要求 $a_{11} \neq 0$，而且要求每经一次消元后得到的 a_{kk} 均不能为零。可以证明，这一要求相当于要求矩阵 A 的各阶主子式均不为零。读者不妨以二阶或三阶主子式进行验证。在实际使用时，当 a_{kk} 较小时，即便可以进行计算，由误差分析知道也会产生较大的误差。所以高斯消去法适用于主对角线上的元素按绝对值比其他元素较大的情况。

二、方法步骤

由前述正逆过程的讨论可知，被消去的元素总归是零，在计算中已不起任何作用，所以在实际描述算法与编制计算程序时，根本不必考虑它们，真正要计算的是每次消元后要保留下来的元素。这样，若计算时将 X 的各分量存于 A_{in+1} $(i = 1, 2, \cdots, n)$ 中，高斯消去法的计算步骤可写为：

(1)消元计算。对于 $k=1,2,\cdots,n-1$,

$$a_{kj}=a_{kj}/a_{kk}\quad (j=k+1,k+2,\cdots,n+1)$$

$$a_{ij}=a_{ij}-a_{ik}a_{kj}\quad (i=k+1,\cdots,n;j=k+1,\cdots,n+1)$$

(2)回代求解。$a_{n\,n+1}=a_{n\,n+1}/a_{nn}$;对于 $k=n-1,n-2,\cdots,1$,

$$a_{k\,n+1}=a_{k\,n+1}-\sum_{j=k+1}^{n}a_{kj}a_{j\,n+1}$$

(3)输出解。$X=(a_{1\,n+1},a_{2\,n+1},\cdots,a_{n\,n+1})^{\mathrm{T}}$。

例 4-1 对乙苯和二甲苯的混合物进行光度分析,在光程长度为 1 cm 时,测得摩尔吸收率数据如表 4-1 所示,试确定混合物中各组分的浓度。

表 4-1

λ/μm	对二甲苯	间二甲苯	邻二甲苯	乙苯	总吸收度 D_λ
12.5	1.502	0.051 4	0	0.040 8	0.101 3
13.0	0.026 1	1.151 6	0	0.082 0	0.099 43
13.4	0.034 2	0.035 5	2.532	0.293 3	0.219 4
14.3	0.034 0	0.068 4	0	0.347 0	0.033 96

解 对于多组分混合物,按照比尔定律有

$$D_{\lambda i}=\sum_{j=1}^{4}m_{ij}c_j\quad (i=1,2,3,4)$$

其中,$D_{\lambda i}$为在波长为 λ_i 时测得的总吸收率;m_{ij}为在波长为 λ_i 时第 j 组分的摩尔吸收率;c_j 为混合物中第 j 组分的摩尔浓度。

由题意可建立线性方程组的矩阵形式为

$$\begin{bmatrix}1.502 & 0.051\,4 & 0 & 0.040\,8\\ 0.026\,1 & 1.151\,6 & 0 & 0.082\,0\\ 0.034\,2 & 0.035\,5 & 2.532 & 0.293\,3\\ 0.034\,0 & 0.068\,4 & 0 & 0.347\,0\end{bmatrix}\begin{bmatrix}c_1\\c_2\\c_3\\c_4\end{bmatrix}=\begin{bmatrix}0.101\,3\\0.099\,43\\0.219\,4\\0.033\,96\end{bmatrix}$$

利用高斯消去法解得

$$c_1=0.062\,7(\mathrm{mol/L}),\quad c_2=0.079\,5(\mathrm{mol/L}),$$

$$c_3=0.075\,9(\mathrm{mol/L}),\quad c_4=0.076\,2(\mathrm{mol/L})$$

第二节　高斯主元素消去法

一、方法的基本思想

为了克服高斯消去法的局限性和提高计算的精度，这里讨论高斯主元素消去法。主元素消去法有多种，在本节只讨论全主元素消去法，其他常用的还有列主元素消去法。

高斯主元素消去法的实施步骤是：在每次消元之前，从当前矩阵中选取一个绝对值最大的元素，即所谓的主元素，并把主元素所在的行叫主行，所在的列叫主列，当第一次消元时，以主行为基础，经过初等变换，消去 n 阶矩阵主列中除主元素之外的其他元素，经过第一次消元后得到一个除去主行主列的元素组成的 $(n-1)$ 阶矩阵，与其对应的为一个 $(n-1)$ 阶的方程组，再以同样的方式对当前的系数矩阵找出主元素并消元，重复以上过程 n 次后，把所有的主行、主列合在一起，便得到一个在不同行不同列上只有一个元素的 n 阶矩阵组成的与原方程组等价的方程组，于是便可直接得到方程组的解。

为了便于理解，下面以一个 4×4 的矩阵给出选取主元素和消元的变化过程。在演变的过程中，仍用动态的表示法，即在元素的相应位置上均使用同一个符号。

$$\left[\begin{array}{cccc:c} a_{11} & a_{12} & \boxed{a_{13}} & a_{14} & b_1 \\ a_{21} & a_{22} & a_{23} & a_{24} & b_2 \\ a_{31} & a_{32} & a_{33} & a_{34} & b_3 \\ a_{41} & a_{42} & a_{43} & a_{44} & b_4 \end{array}\right] \xrightarrow[\text{消元}]{\text{主元:}a_{13}}$$

$$\left[\begin{array}{cccc:c} a_{11} & a_{12} & \boxed{a_{13}} & a_{14} & a_{15} \\ a_{21} & a_{22} & 0 & a_{24} & a_{25} \\ a_{31} & a_{32} & 0 & a_{34} & a_{35} \\ \boxed{a_{41}} & a_{42} & 0 & a_{44} & a_{45} \end{array}\right] \xrightarrow[\text{换行}]{\text{主元:}a_{41}}$$

$$\left[\begin{array}{cccc:c} a_{11} & a_{12} & \boxed{a_{13}} & a_{14} & a_{15} \\ \boxed{a_{21}} & a_{22} & 0 & a_{24} & a_{25} \\ a_{31} & a_{32} & 0 & a_{34} & a_{35} \\ a_{41} & a_{42} & 0 & a_{44} & a_{45} \end{array}\right] \xrightarrow[\text{消元}]{\text{主元:}a_{21}}$$

$$\left[\begin{array}{cccc:c} 0 & a_{12} & \boxed{a_{13}} & a_{14} & a_{15} \\ \boxed{a_{21}} & a_{22} & 0 & a_{24} & a_{25} \\ 0 & a_{32} & 0 & \boxed{a_{34}} & a_{35} \\ 0 & a_{42} & 0 & a_{44} & a_{45} \end{array}\right] \xrightarrow[\text{消元}]{\text{主元:}a_{34}}$$

$$\left[\begin{array}{cccc:c} 0 & a_{12} & \boxed{a_{13}} & 0 & a_{15} \\ \boxed{a_{21}} & a_{22} & 0 & 0 & a_{25} \\ 0 & a_{32} & 0 & \boxed{a_{34}} & a_{35} \\ 0 & \boxed{a_{42}} & 0 & 0 & a_{45} \end{array}\right] \xrightarrow[\text{消元}]{\text{主元}:a_{42}}$$

$$\left[\begin{array}{cccc:c} 0 & & \boxed{a_{13}} & 0 & a_{15} \\ \boxed{a_{21}} & 0 & 0 & 0 & a_{25} \\ 0 & 0 & 0 & \boxed{a_{34}} & a_{35} \\ 0 & \boxed{a_{42}} & 0 & 0 & a_{45} \end{array}\right]$$

在变换过程中，若主元素不在前边行里时，可把它调到较前边，这只是为了编程的方便，也可以不作这种调换直接消元。从以上所得最终结果来看，方程组的解已显而易见。若再注意到主元素所在的列与向量 X 各分量的下标是一致的，在选主元素时记下这些列数，最后就可以把解按顺序输出来。

一般来说，如图 4－1 所示，若主元素在 p 行、q 列，即主元素为 a_{pq}，则第 j 行、第 k 列的元素 a_{jk} 可按下式计算：

$$S=\frac{a_{jq}}{a_{pq}}$$

$$a_{jk}=a_{jk}-a_{pk}S$$

其中，$i,p,q=1,2,\cdots,n;k=1,2,\cdots,n+1$。

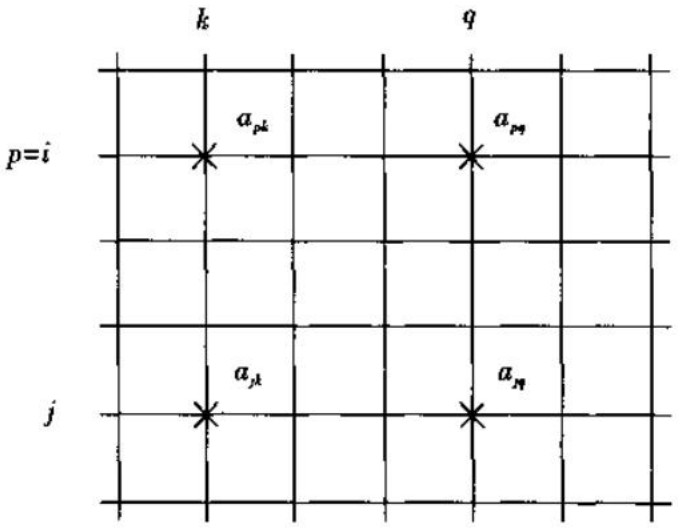


图 4－1

在全部主元素选出并完成全部消元之后，可以按下式计算出未知向量的各个分量[第 i 行的分量为 $X(q)$]：

$$X(q)=a_{i,n+1}/a_{iq}$$

其中，i 为主元素所在行；q 为主元素所在列，这个列标也就是各分量的下标。

二、方法步骤

综上所述，高斯主元素消去法的步骤为：

(1)找主元素，即找出当前矩阵中绝对值最大的元素。

(2)进行换行，即若主元素不在前边行里时，可把它调到较前边。但换行时要注意包括右端项一起交换。

(3)消元计算，即利用主元素所在的行，消去主列中的其他元素。

(4)在选出全部主元素和全部消元结束后，得到一个除主行、主列元素之外，其他元素均为零的矩阵，再依照主元素的列标求得未知量的各分量，即 $X(q)=a_{i,n+1}/a_{iq}$。

(5)输出 $X=(x_1,x_2,\cdots,x_n)^T$。

例 4-2　利用高斯主元素消去法求下面方程组的解：

$$\begin{cases}0.5x_1+1.1x_2+3.1x_3=6.0\\2.0x_1+4.5x_2+0.36x_3=0.02\\5.0x_1+0.96x_2+6.5x_3=0.96\end{cases}$$

解　利用高斯主元素消去法的求解过程为

$$\left[\begin{array}{ccc|c}0.5&1.1&3.1&6.0\\2.0&4.5&0.36&0.02\\5.0&0.96&6.5&0.96\end{array}\right]\xrightarrow[\text{换行}]{\text{主元素：}a_{33}=6.5}$$

$$\left[\begin{array}{ccc|c}5.0&0.96&6.5&0.96\\2.0&4.5&0.36&0.02\\0.5&1.1&3.1&6.0\end{array}\right]\xrightarrow[\text{消元}]{\text{主元素：}a_{13}=6.5}$$

$$\left[\begin{array}{ccc|c}5.0&0.96&6.5&0.96\\1.72&4.45&0&-0.033\\-1.88&0.64&0&5.54\end{array}\right]\xrightarrow[\text{消元}]{\text{主元素：}a_{22}=4.45}$$

$$\left[\begin{array}{ccc|c}4.63&0&6.5&0.97\\1.72&4.45&0&-0.033\\-2.13&0&0&5.54\end{array}\right]\xrightarrow[\text{消元}]{\text{主元素：}a_{31}=-2.13}$$

$$\left[\begin{array}{ccc|c}0&0&6.5&13.01\\0&4.45&0&4.44\\-2.13&0&0&5.54\end{array}\right]$$

则

$$x_1 = -2.6, \quad x_2 = 1.00, \quad x_3 = 2.00$$

第三节　追赶法

当从实际问题形成线性代数方程组时，其系数矩阵往往有两种情况：一种是低阶稠密矩阵，另一种是高阶稀疏矩阵，相应地产生了各种不同的数值解法，这里不去一一介绍。在大型稀疏矩阵中，三对角线矩阵是常见的一种，例如在多元精馏及吸收塔的平衡级模拟计算、三次样条插值以及用差分法解常微分方程的边值问题等的计算中，都会产生三对角线方程组。因而讨论其求解方法具有重要的实用价值。

一、方法介绍

三对角线方程组具有以下形式：

$$\begin{bmatrix} b_1 & c_1 & & & \\ a_2 & b_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & a_{n-1} & b_{n-1} & c_{n-1} \\ & & & a_n & b_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{n-1} \\ d_n \end{bmatrix}$$

即

$$MX = D$$

在系数矩阵 M 中，非零元素集中在主对角线及其两侧的两条对角线上，其他位置上的元素全为零，三对角线矩阵即由此得名。对于这类矩阵，由于其形状的特殊性，可以用所谓的追赶法求解。

追赶法的实现过程与高斯消去法基本类似。只是由于矩阵的特殊性，使得用追赶法求解不仅可以减少计算次数，而且可以压缩数据的存贮单元。

具体地说，追赶法分为正过程和逆过程两部分。正过程又称为追过程，它是通过消元把下对角线的元素 $a_i(i=2,3,\cdots,n)$ 化为零，同时把主对角线上的元素化为 1。变化之后，其增广矩阵具有如下的形式（这里仍用原来的符号）：

$$\left[\begin{array}{ccccc|c} 1 & c_1 & & & & d_1 \\ & 1 & c_2 & & & d_2 \\ & & \ddots & \ddots & & \vdots \\ & & & 1 & c_{n+1} & d_{n-1} \\ & & & & 1 & d_n \end{array}\right]$$

从上式的最后一行可以看出 $x_n = d_n$。若从 x_n 的值出发，把求得的解（分量）逐次回代上边的方程，则可以得到全部分量 $x_i(i=n-1,\cdots,1)$，这个过程称为赶过程。

下边按照三对角矩阵的特点推导消元过程的计算式，同时也给出回代过程的计算

公式。

从以上的讨论不难看出，消元变换过程实际要计算的量只是 c_i 和 $d_i(i=1,2,\cdots,n)$，对于 c_1 和 d_1，显然有

$$c_1 = c_1/b_1, \quad d_1 = d_1/b_1$$

这种消元的特点是，每次只消去一个元素，所以每次计算也限于一行的计算，下边从第二行开始，逐行消去 a_i 并计算 c_i 和 $d_i(i=2,3,\cdots,n)$。

当消去 a_2 时，新的 b_2 应当由原来的 b_2 减去 c_1 乘 a_2 得到，即 $b_2 = b_2 - c_1a_2$。

令 $r = b_2 = b_2 - c_1a_2$，当进一步把 b_2 变为 1 时，c_2 和 d_2 应进行如下的计算：

$$c_2 = c_2/r(\text{即 } c_2 = c_2/b_2), \quad d_2 = (d_2 - d_1a_2)/r$$

依次类推，当从第 k 行消去 a_k 时，若令 $r = b_k - a_kc_{k-1}$，则 c_k, d_k 应按下式计算：

$$c_k = c_k/r, \quad d_k = (d_k - d_{k-1}a_k)/r$$

其中，$k=2,3,\cdots,n$。应注意当 $k=n$ 时，c_n 已不属于矩阵元素了，只有 d_n 是有用的量，这时的 d_n 也正是要求的 x_n。

由于 x_n 已求出，所以其他量的计算可按下式推得：

$$x_k = d_k - c_kx_{k+1} \quad (k=n-1,n-2,\cdots,1)$$

二、计算步骤

在进行追赶法的计算机编程计算时，完全没有必要设置一个数组 $x(n)$ 来存放解 $X=(x_1,\cdots,x_2,x_n)^T$，可直接把计算结果存放于 $d_i(i=1,2,\cdots,n)$ 中，这样可以节省存贮单元。于是追赶法的计算步骤可写为：

(1)输入 $a_i, b_i, c_i, d_i(i=1,2,\cdots,n)$，但需要注意 a_1 和 c_n 不是矩阵的元素；

(2)追赶过程，即消元计算

$$c_1 = c_1/b_1, \quad d_1 = d_1/b_1$$

对于 $k=2,3,\cdots,n$，

$$r = b_k - a_kc_{k-1}, \quad c_k = c_k/r, \quad d_k = (d_k - d_{k-1}a_k)/r;$$

(3)赶过程，即回代求解

$$d_k = (d_k - c_kd_{k+1}) \quad (k=n-1,\cdots,1);$$

(4)输出 $X=(d_1,d_2,\cdots,d_n)^T$。

由此可见，整个计算过程是一个用单下标进行递推的过程，这就可以用一维数组来实现存贮，而避免使用二维存贮，从而大大地压缩了存贮空间。例如，对一个 100 阶的方阵来说，二维存贮要用 10 000 个数据单元，而一维存贮只要用 300 个单元就足够了，随着 n 的增大，这种差别就更加明显。因此，对于三对角矩阵来说使用追赶法不仅提高了计算速度，而且也节省了存贮空间。

例 4-3 用固定床反应器的拟均相二维模型求解乙苯脱氢反应器中沿径向反应物浓度分布时，得到一组有关沿反应器内径向 6 个点处乙苯转化率 x 的方程组：

$$\begin{cases} x_1 - 0.333x_2 = 0.029\,6 \\ 0.05x_1 - x_2 + 0.15x_3 = -0.035\,6 \\ 0.075x_2 - x_3 + 0.125x_4 = -0.035\,6 \\ 0.083\,3x_3 - x_4 + 0.116\,7x_5 = -0.035\,6 \\ 0.087\,5x_4 - x_5 + 0.112\,5x_6 = -0.035\,6 \\ 0.20x_5 - x_6 = -0.047\,2 \end{cases}$$

试求沿反应器内径向上各点转化率。

解 将方程组写成矩阵形式：

$$\begin{bmatrix} 1.0 & -0.333 & & & & \\ 0.05 & -1.0 & 0.15 & & & \\ & 0.075 & -1.0 & 0.125 & & \\ & & 0.083\,3 & -1.0 & 0.116\,7 & \\ & & & 0.087\,5 & -1.0 & 0.112\,5 \\ & & & & 0.20 & -1.0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0.029\,6 \\ -0.035\,6 \\ -0.035\,6 \\ -0.035\,6 \\ -0.035\,6 \\ -0.047\,2 \end{bmatrix}$$

利用追赶法解得

$$\begin{aligned} X &= (x_1, x_2, x_3, x_4, x_5, x_6)^{\mathrm{T}} \\ &= (0.044\,4, 0.044\,5, 0.044\,5, 0.044\,7, 0.044\,8, 0.056\,4)^{\mathrm{T}} \end{aligned}$$

显然，反应器内径向各点处转化率有所不同，这是由于固定床反应器径向温度分布不均匀所致。

第四节 *LU* 分解法

对于 n 阶方阵 A，若存在 n 阶下三角矩阵 L 和 n 阶上三角矩阵 U，使得 $A = LU$，则称 L、U 为矩阵 A 的 LU 分解，亦称为三角分解。一般来说，LU 分解不是唯一的。若矩阵 A 的各阶顺序主子式不为零，即

$$a_{11} \neq 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0, \quad \cdots, \quad \det A \neq 0$$

则存在两种确定的分解形式。

一种是规定 L 是单位下三角矩阵，U 是非奇异的上三角矩阵，即

$$L = \begin{bmatrix} 1 & & & \\ l_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & 1 \end{bmatrix}, \quad U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & u_{nn} \end{bmatrix}$$

这种情况称为杜特利尔分解。

另一种是规定 U 是单位上三角矩阵，L 是非奇异的下三角矩阵，即

$$U=\begin{bmatrix}1 & u_{12} & \cdots & u_{1n}\\ & 1 & \cdots & u_{2n}\\ & & \ddots & \vdots\\ & & & 1\end{bmatrix},\quad L=\begin{bmatrix}l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & l_{nn}\end{bmatrix}$$

这种情况称为克劳特分解。

现对方程组 $AX=B$ 进行 LU 分解，由于 $A=LU$，于是得

$$LUX=B$$

令 $Y=UX$，即把方程组 $AX=B$ 的求解化为求解下述方程组：

$$\begin{cases}LY=B, & \text{求 } Y\\ UX=Y, & \text{求 } X\end{cases}$$

这两个方程组的系数矩阵分别为下三角矩阵和上三角矩阵，很容易求解。这样只对 A 作一次分解，对每个方程组按 $LY=B$ 和 $UX=Y$ 作 n^2 次乘除运算便可求得所有解。

这里以杜特利尔分解为例，讨论如何由 A 的元素 a_{ij} 确定 L 和 U 的元素。依照矩阵乘法有

$$\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}=\begin{bmatrix}1 & & & \\ l_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & 1\end{bmatrix}\begin{bmatrix}u_{11} & u_{12} & \cdots & u_{1n}\\ & u_{22} & \cdots & u_{2n}\\ & & \ddots & \vdots\\ & & & u_{nn}\end{bmatrix}$$

$$=\begin{bmatrix}u_{11} & u_{22} & \cdots & u_{1n}\\ l_{21}u_{11} & l_{21}u_{12}+u_{22} & \cdots & l_{21}u_{1n}+u_{2n}\\ \vdots & \vdots & & \vdots\\ l_{n1}u_{11} & l_{n1}u_{12}+l_{n2}u_{22} & \cdots & \sum_{i=1}^{n}l_{ni}u_{in}\end{bmatrix}$$

即

$$a_{ij}=\sum_{r=1}^{n}l_{ir}u_{rj}\quad(i,j=1,2,\cdots,n)$$

而

$$l_{ij}=\begin{cases}0 & (i<j)\\ 1 & (i=j)\\ l_{ij} & (i>j)\end{cases},\qquad u_{ij}=\begin{cases}u_{ij} & (i<j)\\ u_{ii} & (i=j)\\ 0 & (i>j)\end{cases}$$

则有

$$a_{1j}=u_{1j}\quad(j=1,2,\cdots,n),$$
$$a_{i1}=l_{i1}u_{11}\quad(i=2,3,\cdots,n)$$

对于 $k=2,3,\cdots,n$，有

$$a_{kj} = u_{kj} + \sum_{r=1}^{k-1} l_{kr}u_{rj} \quad (j = k,k+1,\cdots,n),$$

$$a_{ik} = l_{ik}u_{kk} + \sum_{r=1}^{k-1} l_{ir}u_{rk} \quad (i = k+1,\cdots,n,\text{且 } k \neq n)$$

于是有

$$u_{1j} = a_{1j} \quad (j=1,2,\cdots,n),$$

$$l_{i1} = a_{i1}/u_{11} \quad (i=2,3,\cdots,n)$$

对于 $k=2,3,\cdots,n$，有

$$u_{kj} = a_{kj} - \sum_{r=1}^{k-1} l_{kr}u_{rj} \quad (j = k,k+1,\cdots,n),$$

$$l_{ik} = (a_{ik} - \sum_{r=1}^{k-1} l_{ir}u_{rk})/u_{kk} \quad (i = k+1,\cdots,n,\text{且 } k \neq n)$$

将 A 分解为 L 和 U 后，则可先由式 $LY=B$ 求 Y：

$$y_1 = b_1,$$

$$y_k = b_k - \sum_{r=1}^{k-1} l_{kr}y_r \quad (k = 2,3,\cdots,n)$$

然后再由式 $UX=Y$ 求 X：

$$x_n = y_n/u_{nn},$$

$$x_k = (y_k - \sum_{r=k+1}^{n} u_{kr}x_r)/u_{kk} \quad (k = n-1,\cdots,1)$$

实际上，式 $LY=B$ 是解单位下三角线性方程组的计算公式，而 $UX=Y$ 是解上三角线性方程组的计算公式。其实，A 的三角分解及求 Y 的过程相当于高斯消去法的消元计算，而求 X 的过程相当于回代。

由于 L 为单位下三角矩阵，且 $l_{kk}=1(k=1,2,\cdots,n)$，在求解过程中不必存贮，故 LU 分解计算可在矩阵 A 的位置上进行，在主对角线及上三角部分存放 U 矩阵各元素；在下三角部分存放 L 矩阵各元素。若再令 $A_{n+1}=B$，且将 Y 和 X 先后储存其中，那么用 LU 分解法求解线性方程组的计算步骤可写为：

(1)进行 LU 分解

1)计算 L 的第一列元素 $a_{i1}=a_{i1}/a_{11}(i=2,\cdots,n)$；

2)计算 U 的第 k 行元素($k=2,3,\cdots,n$)

$$a_{kj} = a_{kj} - \sum_{r=1}^{k-1} a_{kr}a_{rj}(j = k,k+1,\cdots,n);$$

3)计算 L 的第 k 列元素($k=2,3,\cdots,n$)

$$a_{ik} = (a_{ik} - \sum_{r=1}^{k-1} a_{ir}a_{rk})/a_{kk} \quad (i = k+1,\cdots,n,\text{且 } k \neq n);$$

(2)解 $LY=A_{n+1}$ 求 $y_k(y_1=a_{1n+1})$

$$a_{k\,n+1} = a_{k\,n+1} - \sum_{r=1}^{k-1} a_{kr}a_{r\,n+1} \quad (k = 2,3,\cdots,n);$$

(3)解 $UX = Y$ 求 x_k

$$a_{n\,n+1} = a_{n\,n+1}/a_{nn},$$

$$a_{k\,n+1} = (a_{k\,n+1} - \sum_{r=k+1}^{n} a_{kr}a_{r\,n+1})/a_{kk} \quad (k = n-1,\cdots,1);$$

(4)输出解 $X = A_{n+1}$。

例 4-4 乙炔的摩尔热容与温度的经验关系式为 $C_p = a + bT + cT^2$，用最小二乘法拟合实测数据得正则方程组

$$\begin{cases} 8a + 28b + 140c = 105.21 \\ 28a + 140b + 784c = 402.29 \\ 140a + 784b + 4\,676c = 2\,070.29 \end{cases}$$

试用 LU 分解法求解方程组，确定参数 a、b、c。

解 首先将方程组的系数矩阵 A 作 LU 分解：

$$A = \begin{bmatrix} 8 & 28 & 140 \\ 28 & 140 & 784 \\ 140 & 784 & 4\,676 \end{bmatrix}, \quad L = \begin{bmatrix} 1 & 0 & 0 \\ 3.5 & 1 & 0 \\ 17.5 & 7 & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 8 & 28 & 140 \\ 0 & 42 & 294 \\ 0 & 0 & 168 \end{bmatrix}$$

然后由 $LY = B$ 求 $y_k(k = 1,2,3)$，即

$$\begin{bmatrix} 1 & 0 & 0 \\ 3.5 & 1 & 0 \\ 17.5 & 7 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 105.21 \\ 402.29 \\ 2\,070.29 \end{bmatrix}$$

解得

$$y_1 = 105.21, \quad y_2 = 34.055, \quad y_3 = -9.27$$

再由 $UX = Y$ 求 a、b、c，即

$$\begin{bmatrix} 8 & 28 & 140 \\ 0 & 42 & 294 \\ 0 & 0 & 168 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 105.21 \\ 34.055 \\ -9.27 \end{bmatrix}$$

求得

$$a = 9.927\,1, \quad b = 1.197\,2, \quad c = -0.055\,2$$

第五节 LDL^T 分解法

一、LDL^T 分解法

当 $AX = B$ 中的矩阵 A 非奇异且对称，即 $\det A \neq 0$ 且 $A = A^T$ 时，将 A 分解为 L 和 U 后，

U 尚可进一步分解，即 $A = LU = LDU_0$。其中，L 为单位下三角矩阵，U_0 为单位上三角矩阵，D 为对角矩阵。其展开式为

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} 1 & & & \\ l_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} u_{11} & & & \\ & u_{22} & & \\ & & \ddots & \\ & & & u_{nn} \end{bmatrix} \begin{bmatrix} 1 & u_{12}/u_{11} & \cdots & u_{1n}/u_{11} \\ & 1 & \cdots & u_{2n}/u_{22} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}$$

由于 A 对称，故有

$$A = A^{\mathrm{T}} = (LDU_0)^{\mathrm{T}} = U_0^{\mathrm{T}} D L^{\mathrm{T}}$$

于是

$$L = U_0^{\mathrm{T}}, \quad U_0 = L^{\mathrm{T}}, \quad A = LDL^{\mathrm{T}}$$

对矩阵 A 作的这样的分解法称为 LDL^{T} 分解法。

事实上，只要 A 对称且非奇异，则 A 可唯一地分解为 LDL^{T}。因此有

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} 1 & & & \\ l_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix} \begin{bmatrix} 1 & l_{21} & \cdots & l_{n1} \\ & 1 & \cdots & l_{n2} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}$$

这样，只要求得 L 和 D 的各元素，再由 L 转置便可得 L^{T} 各元素，使得 A 的分解计算几乎减半。

按照矩阵乘法，对于下三角部分有

$$a_{11} = d_1,$$

$$a_{ij} = \sum_{k=1}^{n} l_{ik} d_k l_{jk} = \sum_{k=1}^{j-1} l_{ik} d_k l_{jk} + l_{ij} d_j \quad (\text{当 } j < k \text{ 时}, l_{jk} = 0)$$

因此，可得 L 和 D 矩阵各元素

$$d_1 = a_{11},$$

$$l_{i1} = a_{i1}/d_1 \quad (i = 2, 3, \cdots, n)$$

对于 $j = 2, 3, \cdots, n$，有

$$d_j = a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2 d_k,$$

$$l_{ij} = (a_{ij} - \sum_{k=1}^{j-1} l_{jk} d_k l_{ik})/d_j \quad (i = j+1, \cdots, n; j \neq n)$$

将 $A = LDL^{\mathrm{T}}$ 分解法用于解系数矩阵对称的线性方程组 $AX = B$，则有

$$LDL^{\mathrm{T}} X = B$$

令 $Y = DL^{\mathrm{T}} X$，则

$$LY = B$$

这样，就有

$$y_1 = b_1,$$

$$y_j = b_j - \sum_{k=1}^{j-1} l_{jk} y_k \quad (j = 2,3,\cdots,n);$$

$$x_n = y_n / d_n,$$

$$x_i = y_i / d_i - \sum_{k=i+1}^{n} l_{ki} x_k \quad (i = n-1,\cdots,1)$$

于是，LDL^T 分解法的计算步骤为：

(1)计算 L 的第一列元素

$$a_{i1} = a_{i1} / a_{11} \quad (i = 2,3\cdots,n);$$

(2)在下三角部分计算 D 和 L(第一列元素除外)

对于 $j = 2,3,\cdots,n$，有

$$a_{jj} = a_{jj} - \sum_{k=1}^{j-1} a_{jk}^2 a_{kk},$$

$$a_{ij} = (a_{ij} - \sum_{k=1}^{j-1} a_{ik} a_{jk} a_{kk}) / a_{jj} \quad (i = j+1,\cdots,n, j \neq n);$$

(3)解 $LY = A_{n+1}$ 求 y_j

$$a_{jn+1} = a_{jn+1} - \sum_{k=1}^{j-1} a_{jk} a_{kn+1} \quad (j = 2,3,\cdots,n);$$

(4)解 $DL^TX = Y$ 求 x_i

$$a_{nn+1} = a_{nn+1} / a_{nn},$$

$$a_{in+1} = a_{in+1} / a_{ii} - \sum_{k=i+1}^{n} a_{ki} a_{kn+1} \quad (i = n-1,\cdots,1);$$

(5)输出解 $X = A_{n+1} = (a_{1n+1}, a_{2n+1}, \cdots, a_{nn+1})^T$。

显然，由于 A 对称，分解仅在下三角部分进行，且将分解结果也存放于 A 的下三角部分，使得计算工作量和贮存单元可大约节省一半。再将中间结果 Y 和解 X 存放于列向量 A_{n+1} 中，又可节省 n 个储存单元。

例 4-5 利用 LDL^T 分解法求解下列系数矩阵对称的线性方程组：

$$\begin{bmatrix} 5 & 7 & 6 & 5 & 1 \\ 7 & 10 & 8 & 7 & 2 \\ 6 & 8 & 10 & 9 & 3 \\ 5 & 7 & 9 & 10 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 24 \\ 34 \\ 36 \\ 35 \\ 15 \end{bmatrix}$$

解 由于系数矩阵 A 对称，所以在输入数据时只输入下三角部分即可。L 和 D 各元素为

$$L=\begin{bmatrix}1&0&0&0&0\\1.4&1&0&0&0\\1.2&-2&1&0&0\\1.0&0&1.5&1&0\\0.2&3&1.5&-3&1\end{bmatrix},\quad D=\begin{bmatrix}5&0&0&0&0\\0&0.2&0&0&0\\0&0&2&0&0\\0&0&0&0.5&0\\0&0&0&0&-6\end{bmatrix}$$

于是由

$$\begin{bmatrix}1&0&0&0&0\\1.4&1&0&0&0\\1.2&-2&1&0&0\\1.0&0&1.5&1&0\\0.2&3&1.5&-3&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\\y_4\\y_5\end{bmatrix}=\begin{bmatrix}24\\34\\36\\35\\15\end{bmatrix}$$

求得

$$y_1=24,\quad y_2=0.4,\quad y_3=8,\quad y_4=-1,\quad y_5=-6$$

由

$$\begin{bmatrix}5&0&0&0&0\\0&0.2&0&0&0\\0&0&2&0&0\\0&0&0&0.5&0\\0&0&0&0&-6\end{bmatrix}\begin{bmatrix}1&1.4&1.2&1.0&0.2\\0&1&-2&0&3\\0&0&1&1.5&1.5\\0&0&0&1&-3\\0&0&0&0&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\end{bmatrix}=\begin{bmatrix}24\\0.4\\8\\-1\\-6\end{bmatrix}$$

求得

$$x_1=1,\quad x_2=1,\quad x_3=1,\quad x_4=1,\quad x_5=1$$

二、平方根法

当方程组 $AX=B$ 的系数矩阵 A 为对称正定矩阵时，其 LDL^{T} 分解还可以进一步简化。由前述可知，若矩阵 A 为 n 阶非奇异的对称矩阵，则它可唯一地分解为 $A=LDL^{\mathrm{T}}$，其中 L 为单位下三角矩阵，D 为非奇异对角矩阵。

现因 A 为正定矩阵，则 D 的元素皆为正数，于是 D 又可分解为

$$\begin{bmatrix}d_1&&&\\&d_2&&\\&&\ddots&\\&&&d_n\end{bmatrix}=\begin{bmatrix}\sqrt{d_1}&&&\\&\sqrt{d_2}&&\\&&\ddots&\\&&&\sqrt{d_n}\end{bmatrix}\begin{bmatrix}\sqrt{d_1}&&&\\&\sqrt{d_2}&&\\&&\ddots&\\&&&\sqrt{d_n}\end{bmatrix}$$

这样

$$A=LDL^{\mathrm{T}}=LD^{\frac{1}{2}}D^{\frac{1}{2}}L^{\mathrm{T}}=(LD^{\frac{1}{2}})(LD^{\frac{1}{2}})^{\mathrm{T}}=L_1L_1^{\mathrm{T}}$$

对称正定矩阵的三角分解 $A=L_1L_1^{\mathrm{T}}$ 又称为乔累斯基分解，它可具体写为

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & \cdots & l_{n1} \\ & l_{22} & \cdots & l_{2n} \\ & & \ddots & \vdots \\ & & & l_{nn} \end{bmatrix}$$

这样，只要求得矩阵 L_1 的全部元素，L_1^{T} 也就确定了。其计算只是 LU 分解法的二分之一。

由矩阵的乘法，对于下三角部分

$$a_{11} = l_{11}^2, \quad a_{i1} = l_{i1} l_{11} \quad (i = 2,3,\cdots,n)$$

对于 $j = 2,3,\cdots,n$，有

$$a_{ij} = \sum_{k=1}^{n} l_{ik} l_{jk} = \sum_{k=1}^{j-1} l_{ik} l_{jk} + l_{ij} l_{jj}$$

其中，当 $k > j$ 时，$l_{jk} = 0$；$i = j+1,\cdots,n$；$j \neq n$。

于是

$$l_{11} = \sqrt{a_{11}}, \quad l_{i1} = a_{i1} / l_{11} \quad (i = 2,3,\cdots,n)$$

对于 $j = 2,3,\cdots,n$，有

$$l_{jj} = \left[a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2 \right]^{\frac{1}{2}},$$

$$l_{ij} = \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk} \right) / l_{jj} \quad (i = j+1,\cdots,n; j \neq n)$$

利用乔累斯基分解求解方程组的方法也称为平方根法。对于方程组 $AX = B$，若 A 为对称正定矩阵，则由乔累斯基分解得 $A = L_1 L_1^{\mathrm{T}}$，于是方程组可写为

$$L_1 L_1^{\mathrm{T}} X = B$$

令 $L_1^{\mathrm{T}} X = Y$，于是

$$L_1 Y = B$$

这样首先利用 $L_1 Y = B$ 求 y_k：

$$y_1 = b_1 / l_{11},$$

$$y_k = \left(b_K - \sum_{r=1}^{k-1} l_{kr} y_r \right) / l_{kk} \quad (k = 2,3,\cdots,n)$$

再由 $L_1^{\mathrm{T}} X = Y$ 求 x_k：

$$x_n = y_n / l_{nn},$$

$$x_k = \left(y_k - \sum_{r=k+1}^{n} l_{rk} x_r \right) / l_{kk} \quad (k = n-1,\cdots,1)$$

平方根法的计算步骤为：

(1)计算 L_1 的第一列元素

$$a_{11} = \sqrt{a_{11}}, \quad a_{i1} = a_{i1} / a_{11} \quad (i = 2,3,\cdots,n);$$

(2)在下三角部分计算 L_1(第一列元素除外)

$$a_{jj} = \left(a_{jj} - \sum_{k=1}^{j-1} a_{jk}^2\right)^{\frac{1}{2}},$$

$$a_{ij} = \left(a_{ij} - \sum_{k=1}^{j-1} a_{ik}a_{jk}\right)/a_{jj} \quad (i = j+1,\cdots,n,\text{且}j \neq n);$$

(3)解 $L_1 Y = A_{n+1}$ 求 y_k

$$a_{1\,n+1} = a_{1\,n+1}/a_{11},$$

$$a_{k\,n+1} = \left(a_{k\,n+1} - \sum_{r=1}^{k-1} a_{kr}a_{r\,n+1}\right)/a_{kk} \quad (k = 2,3,\cdots,n);$$

(4)解 $L_1^{\mathrm{T}} X = Y$ 求 x_k

$$a_{n\,n+1} = a_{n\,n+1}/a_{nn},$$

$$a_{k\,n+1} = \left(a_{k\,n+1} - \sum_{r=k+1}^{n} a_{rk}a_{r\,n+1}\right)/a_{kk} \quad (k = n-1,\cdots,1);$$

(5)输出解 $X = A_{n+1}$。

例 4-6 试用乔累斯基分解求解下列方程组:

$$\begin{bmatrix} 5 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 92 \\ 128 \\ 132 \\ 124 \end{bmatrix}$$

解 系数矩阵 A 为对称正定矩阵,将其进行乔累斯基分解:

$$L_1 = \begin{bmatrix} 2.236 & 0 & 0 & 0 \\ 3.131 & 0.444 & 0 & 0 \\ 2.683 & -0.902 & 1.410 & 0 \\ 2.236 & -0.002\,06 & 2.127 & 0.690 \end{bmatrix}$$

由

$$\begin{bmatrix} 2.236 & 0 & 0 & 0 \\ 3.131 & 0.444 & 0 & 0 \\ 2.683 & -0.902 & 1.410 & 0 \\ 2.236 & -0.002\,06 & 2.127 & 0.690 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 92 \\ 128 \\ 132 \\ 124 \end{bmatrix}$$

解得

$$y_1 = 41.145, \quad y_2 = -1.858, \quad y_3 = 14.136, \quad y_4 = 2.795$$

再由

$$\begin{bmatrix} 2.236 & 3.131 & 2.683 & 2.236 \\ 0 & 0.444 & -0.902 & -0.002\,06 \\ 0 & 0 & 1.410 & 2.127 \\ 0 & 0 & 0 & 0.690 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 41.145 \\ -1.858 \\ 14.136 \\ 2.795 \end{bmatrix}$$

解得

$$x_1=4,\quad x_2=4,\quad x_3=4,\quad x_4=4$$

第六节　向量与矩阵的范数

解线性代数方程组实质上是按照某种方法确定满足方程 $AX=B$ 的向量 X^*。这个向量称为解向量，这从前几节讨论的所谓精确方法中已有所体会。为了讨论解线性代数方程组的逐次逼近法（即迭代法），特别是研究方法的收敛性问题，必须引入向量与矩阵范数的概念。

一、向量逼近与向量的长度

设 $\{X^{(k)}\}$ 是 n 维向量空间 R^n 中的向量序列，向量 $X^*\in R^n$，则当 $\lim\limits_{k\to\infty}X^{(k)}=X^*$ 时，称向量序列 $\{X^{(k)}\}$ 逼近于 X^*。

这里的逼近指的是它们的相应分量有以下关系：

$$\lim_{k\to\infty}x_i^{(k)}=x_i^*\quad(i=1,2,,\cdots,n)$$

从另一个角度来说，向量 $X^{(k)}$ 趋向于向量 X^* 指的是空间中两点距离越来越小，以至于趋于零。

以二维空间为例，两个向量即平面上的两点，如图 4－2 所示，它们的距离就是差向量 $X^{(k)}-X^*$ 的欧几里得长度

$$l=\sqrt{[x_1^{(k)}-x_1^*]^2+[x_2^{(k)}-x_2^*]^2}$$

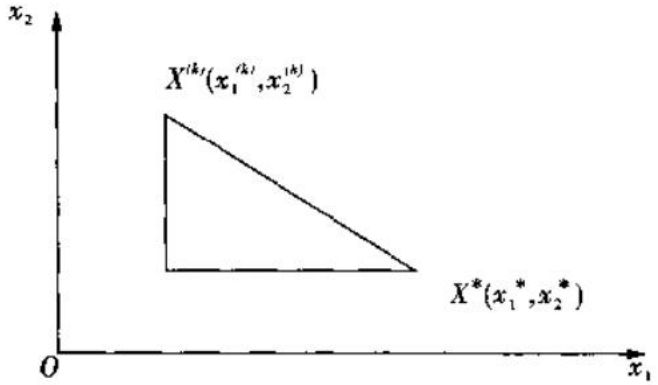


图 4－2

然而，欧几里得长度并不是度量两向量距离远近的唯一方法。例如，可用直角三角形较大的直角边或两直角边之和来度量两点远近的程度，即

$$l=\max\{|x_1^{(k)}-x_1^*|,|x_2^{(k)}-x_2^*|\}\quad 或\quad l=|x_1^{(k)}-x_1^*|+|x_2^{(k)}-x_2^*|$$

若用 X 代表差向量，则向量 X 的长度可用以下 3 种形式之一来描述：

$$l=|x_1|+|x_2|,\quad l=\sqrt{x_1^2+x_2^2},\quad l=\max\{|x_1|,|x_2|\}$$

以上3种形式可直接推广到 n 维向量空间，即

$$l=\sum_{i=1}^{n}|x_i|,\quad l=\sqrt{\sum_{i=1}^{n}x_i^2},\quad l=\max\{|x_1|,|x_2|,\cdots,|x_n|\}$$

那么，究竟什么样的量可以用来反映向量长度的实质，又如何抽象地描述这个量，这些就是要讨论的范数的概念。

二、向量的范数

定义 设向量 $X\in\mathbf{R}^n$，则满足以下3个条件的数 $\|X\|$ 称为向量 X 的范数：

(1) $\|X\|\geqslant 0$，当且仅当 $X=0$ 时等式成立；

(2) 对于任意实数 α，恒有 $\|\alpha X\|=|\alpha|\,\|X\|$；

(3) 对于任何 X、$Y\in\mathbf{R}^n$，成立三角不等式

$$\|X+Y\|\leqslant\|X\|+\|Y\|。$$

不难证明上述的3种向量长度均满足以上3个条件，它们均是一种具体的范数，分别记为

$$\|X\|_1=\sum_{i=1}^{n}|x_i|$$

$$\|X\|_2=\left(\sum_{i=1}^{n}x_i^2\right)^{\frac{1}{2}}$$

$$\|X\|_\infty=\max\{|x_1|,|x_2|,\cdots,|x_n|\}$$

这是常用的向量范数的3种表示形式，依次称为向量 X 的1－范数、2－范数（欧几里得范数）和∞－范数（最大模范数）。它们还可以统一记为如下形式：

$$\|X\|_p=\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}\quad(p=1,2,\infty)$$

$p=1,2$ 的情形是显然的，下面仅对 $p=\infty$ 的情形给予证明。

事实上，$\|X\|_p=\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}$，现记 $|x_k|=\max\{|x_1|,|x_2|,\cdots,|x_n|\}$。

于是

$$|x|_p=\left(\sum_{i=1}^{n}|x_k|^p\left|\frac{x_i}{x_k}\right|^p\right)^{\frac{1}{p}}=|x_k|\left(\sum_{i=1}^{n}\left|\frac{x_i}{x_k}\right|^p\right)^{\frac{1}{p}}$$

而

$$|x_k|^p\leqslant\sum_{i=1}^{n}|x_i|^p\leqslant n|x_k|^p$$

所以

$$1\leqslant\left(\sum_{i=1}^{n}\left|\frac{x_i}{x_k}\right|^p\right)^{\frac{1}{p}}\leqslant n^{\frac{1}{p}}$$

由于

$$\lim_{p\to\infty} n^{\frac{1}{p}} = 1$$

于是

$$\lim_{p\to\infty}\left(\sum_{i=1}^{n}\left|\frac{x_i}{x_k}\right|^p\right)^{\frac{1}{p}} = 1$$

则

$$\|X\|_\infty = \lim_{p\to\infty}\|X\|_p = \lim_{p\to\infty}|x_k|\left(\sum_{i=1}^{n}\left|\frac{x_i}{x_k}\right|^p\right)^{\frac{1}{p}} = |x_k| = \max_i|x_i|$$

由向量范数的定义可以导出下列基本性质：

(1)零向量的范数是零；

(2)当 $X\neq 0$ 时，有 $\left\|\frac{X}{\|X\|}\right\| = \frac{1}{\|X\|}\|X\| = 1$；

(3)对于任意 $X\in R^n$ 有 $\|-X\| = |-1|\,\|X\| = \|X\|$；

(4)对任意的 $X、Y\in R^n$ 有 $\|X-Y\| \geqslant |\,\|X\| - \|Y\|\,|$。

事实上

$$\|X\| = \|X+Y-Y\| \leqslant \|X-Y\| + \|Y\|$$

于是

$$\|X-Y\| \geqslant \|X\| - \|Y\|$$

又

$$\|X-Y\| = \|Y-X\| \geqslant \|Y\| - \|X\|$$

所以

$$\|X\| - \|Y\| \geqslant -\|X-Y\|$$

即

$$-\|X-Y\| \leqslant \|X\| - \|Y\| \leqslant \|X-Y\|$$

所以

$$|\,\|X\| - \|Y\|\,| \leqslant \|X-Y\|$$

凡是按范数定义确定的任何范数均是等价的。也就是说，设 R^n 中的任一向量 X 的两种范数分别为 $\|X\|_\alpha$、$\|X\|_\beta$，则总有两个正数 c_1, c_2 使得成立以下不等式：

$$c_1\|X\|_\beta \leqslant \|X\|_\alpha \leqslant c_2\|X\|_\beta$$

例如：$\|X\|_\infty \leqslant \|X\|_1 \leqslant n\|X\|_\infty$，$\|X\|_\infty \leqslant \|X\|_2 \leqslant \sqrt{n}\|X\|_\infty$。

这是关于向量范数的一个十分重要的性质，由此性质可证明关于向量序列 $\{X^{(k)}\}$ 收敛性的重要定理。

定理 在 R^n 中，向量序列 $\{X^{(k)}\}$ 收敛于向量 X^* 的充分必要条件是

$$\lim_{k\to\infty}\|x^{(k)} - x^*\| = 0$$

换句话说，序列 $\{X^{(k)}\}$ 依任一种范数收敛于 X^*。

证明 由于在 n 维向量空间中各种范数的等价性，所以仅按某一种具体范数加以证明即可。这里采用 2 - 范数进行证明。

充分性：设 $\lim\limits_{k\to\infty}\| X^{(k)}-X^*\|_2=0$，即

$$\lim_{k\to\infty}\left\{\sum_{i=1}^{n}[x_i^{(k)}-x_i^*]^2\right\}^{\frac{1}{2}}=0$$

从而

$$\lim_{k\to\infty}\{[x_i^{(k)}-x_i^*]^2\}^{\frac{1}{2}}=0$$

所以

$$\lim_{k\to\infty}x_i^{(k)}=x_i^* \quad (i=1,2,\cdots,n)$$

于是

$$\lim_{k\to\infty}X^{(k)}=X^*$$

必要性：设 $\lim\limits_{k\to\infty}X^{(k)}=X^*$，则

$$\lim_{k\to\infty}x_i^{(k)}=x_i^* \quad (i=1,2,\cdots,n)$$

从而

$$\lim_{k\to\infty}\left\{\sum_{i=1}^{n}[x_i^{(k)}-x_i^*]^2\right\}^{\frac{1}{2}}=0$$

即

$$\lim_{k\to\infty}\| X^{(k)}-X^*\|_2=0$$

三、矩阵的范数

1. 定义

对任一 n 阶方阵 A，按照一定规则确定一个实数 $\| A\|$ 与之对应，若 $\| A\|$ 满足：

(1) 正定条件，即当 $A\neq 0$ 时，$\| A\|>0$；

(2) 齐次条件，即对任何实数 α，$\| \alpha A\|=|\alpha|\| A\|$；

(3) 三角不等式，即对任意的 n 阶方阵 A、B，有 $\| A+B\|\leqslant\| A\|+\| B\|$；

(4) 对任意的 n 阶方阵 A、B，有 $\| A\cdot B\|\leqslant\| A\|\cdot\| B\|$。

则称实数 $\| A\|$ 为矩阵 A 的范数。

满足上述条件的实数很多，它们都称为矩阵的范数，实际中，常用的矩阵范数有以下 3 种：

(1) 矩阵“F”范数 $\| A\|_F=\left(\sum\limits_{i,j=1}^{n}a_{ij}^2\right)^{\frac{1}{2}}$；

(2) 矩阵行范数 $\| A\|_\infty=\max\limits_{1\leqslant i\leqslant n}\sum\limits_{j=1}^{n}|a_{ij}|$；

(3) 矩阵列范数 $\| A\|_1=\max\limits_{1\leqslant j\leqslant n}\sum\limits_{i=1}^{n}|a_{ij}|$。

2. 矩阵范数在解线性方程组中的作用

虽然矩阵本身通常并不讲什么数量的大小，它只是一些元素的一个表列，然而，正如已经看到的，在解线性方程组时，矩阵要与向量一起运算，不同的矩阵作用在同一个向量上，或同一个矩阵作用在不同的向量上的时候，通常产生出不同的向量。可想而知，这一定与矩阵的元素的数量关系有关。因此，如何从数量的角度来看待矩阵，确实是一个应当研究的问题，其内容构成了矩阵理论的一个重要方面。所谓矩阵范数就是从数量的观点对矩阵作出一种度量。

设有如下形式的线性方程组

$$X = AX + B$$

其中，A 为矩阵，X、B 为向量。

若采用类似于解非线性方程的迭代法对该方程组求解，则可得以下迭代序列：

$$X^{(k+1)} = AX^{(k)} + B \quad (k = 0, 1, 2\cdots)$$

另一方面，如果 X^* 是方程组 $X = AX + B$ 的解，则有

$$X^* = AX^* + B$$

而 $X^{(k)}$ 能否作为 X^* 的近似值问题，要看 $X^{(k)}$ 与 X^* 的差向量 $X^{(k)} - X^*$ 的大小如何而定。而由以上分析可得

$$X^{(k+1)} - X^* = A(X^{(k)} - X^*) \quad (k = 0, 1, 2\cdots)$$

从这个式子可以看出，差向量 $X^{(k+1)} - X^*$ 是由差向量 $X^{(k+1)} - X^*$ 经矩阵 A 的作用后得到的，所以向量 $X^{(k)}$ 是否接近于(收敛于)X^* 的问题，实质上是一个向量 X 经矩阵 A 作用后所得向量 AX 的范数 $\| AX \|$ 是否比原来向量 X 的范数 $\| X \|$ 还小的问题。可见这里唯一起作用的因素就是矩阵 A。

可以证明，$\| AX \| \leqslant \| A \| \| X \|$。这样，若 $\| A \| < 1$，则向量 X 经 A 作用后所得向量 AX 的范数 $\| AX \|$ 比 X 的范数 $\| X \|$ 缩小了 $\| A \|$ 倍。正是具有这样范数的矩阵在方程组求解的迭代法收敛中扮演了主要角色。

第七节　解线性方程组的普通迭代法

一、方法介绍

设 n 阶线性方程组的矩阵形式为

$$MX = G$$

其中，M 为 n 阶方阵，X、G 为 n 维向量。

当用迭代法确定方程组中的向量 X 的近似值时，必须先把它化为便于迭代的形式，即

$$X = AX + B$$

然后选取一个初始向量 $X^{(0)} = (x_1^{(0)}, x_2^{(0)}, \cdots, x_n^{(0)})^T$ 作为解的初始近似值，把 $X^{(0)}$ 代

入式 $X=AX+B$ 的右边，算出解的第一次近似值，记为 $X^{(1)}$。如此反复迭代便得到一个向量序列

$$X^{(1)}=AX^{(0)}+B$$
$$X^{(2)}=AX^{(1)}+B$$
$$\cdots\cdots\cdots\cdots$$
$$X^{(k)}=AX^{(k-1)}+B$$
$$\cdots\cdots\cdots\cdots$$

如果 $X_1^{(0)},X_2^{(1)},\cdots,X_n^{(k)}\cdots$ 的极限存在，即

$$\lim_{k\to\infty}X^{(k)}=X^*$$

那么就称迭代法收敛，其极限 X^* 也就是方程 $X=AX+B$ 的解，从而也就是 $MX=G$ 的解。于是便可以用有限步迭代得到的结果作为 X^* 的近似值。若序列的极限不存在，那么使用迭代法就毫无意义。可见，序列 $\{X^{(k)}\}$ 的收敛与否是使用迭代法的关键，而序列 $\{X^{(k)}\}$ 的收敛与否显然与 $X=AX+B$ 的具体形式有关，或者说与矩阵 A 的具体形式有关。究竟什么样的矩阵 A 能保证迭代收敛，这是要首先弄清的问题。

二、迭代法收敛的充分判别法

定理 若方程组 $X=AX+B$ 中矩阵 A 的某种范数 $\|A\|=q<1$，则对于任何初始向量 $X^{(0)}=(x_1^{(0)},x_2^{(0)},\cdots,x_n^{(0)})^T$，由迭代公式

$$X^{(k+1)}=AX^{(k)}+B\quad(k=0,1,2,\cdots)$$

产生的向量序列 $\{X^{(k)}\}$ 有以下性质：

(1) $\lim_{k\to\infty}X^{(k)}=X^*$；

(2) $\|X^*-X^{(k)}\|\leqslant\frac{q}{1-q}\|X^k-X^{(k-1)}\|$；

(3) $\|X^*-X^{(k)}\|\leqslant\frac{q^k}{1-q}\|X^{(1)}-X^{(0)}\|$。

证明 (1) 设 X^* 为方程组的解，即 $X^*=AX^*+B$，于是

$$X^{(k+1)}-X^*=A(X^{(k)}-X^*)$$

考察它的范数应有

$$\|X^{(k+1)}-X^*\|=\|A(X^{(k)}-X^*)\|\leqslant\|A\|\,\|X^{(k)}-X^*\|$$
$$=q\|X^{(k)}-X^*\|$$

这样，当 $k=0$ 时

$$\|X^{(1)}-X^*\|\leqslant q\|X^{(0)}-X^*\|$$

当 $k=1$ 时

$$\|X^{(2)}-X^*\|\leqslant q\|X^{(1)}-X^*\|\leqslant q^2\|X^{(0)}-X^*\|$$

同理可推得

$$\|X^{(k+1)}-X^*\|\leqslant q^{(k+1)}\|X^{(0)}-X^*\|$$

因为 $0<q<1$,所以

$$\lim_{k\to\infty}\| X^{(k)}-X^{*}\| \leqslant \lim_{k\to\infty} q^{k}\| X^{(0)}-X^{*}\| =0$$

故

$$\lim_{k\to\infty}\| X^{(k)}-X^{*}\| =0$$

即

$$\lim_{k\to\infty} X^{(k)}=X^{*}$$

(2)由于

$$X^{(k)}=AX^{(k+1)}+B,\quad X^{(k+1)}=AX^{(k)}+B$$

于是

$$X^{(k+1)}-X^{(k)}=A(X^{(k)}-X^{(k-1)})$$

则

$$\begin{aligned}\| X^{(k+1)}-X^{(k)}\| &\leqslant \| A\|\ \| X^{(k)}-X^{(k-1)}\| \\ &=q\| X^{(k)}-X^{(k-1)}\| \quad (k=1,2\cdots)\end{aligned}$$

而

$$\begin{aligned}\| X^{(k+1)}-X^{(k)}\| &=\| X^{*}-X^{(k)}-(X^{*}-X^{(k+1)})\| \\ &\geqslant \| X^{*}-X^{(k)}\| -\| X^{*}-X^{(k+1)}\| \\ &\geqslant \| X^{*}-X^{(k)}\| -q\| X^{*}-X^{(k)}\| \\ &=(1-q)\| X^{*}-X^{(k)}\|\end{aligned}$$

所以

$$\| X^{*}-X^{(k)}\| \leqslant \frac{1}{1-q}\| X^{(k+1)}-X^{(k)}\|$$

故

$$\| X^{*}-X^{(k)}\| \leqslant \frac{q}{1-q}\| X^{(k)}-X^{(k-1)}\|$$

(3)由于

$$\| X^{k+1}-X^{(k)}\| \leqslant q\| X^{(k)}-X^{(k-1)}\|$$

结合(2)的结果可立即推出

$$\| X^{*}-X^{(k)}\| \leqslant \frac{q^{k}}{1-q}\| X^{(1)}-X^{(0)}\|$$

由于向量的 1 - 范数、∞ - 范数具有比较简单的形式,使用起来较为方便,所以实践中常用这两种范数来判断收敛性和估计误差。若用这两种具体的范数,则

$$\| X^{*}-X^{(k)}\| \leqslant \frac{q^{k}}{1-q}\| X^{(1)}-X^{(0)}\|$$

有如下形式:

相应于 1 - 范数误差估计式为

$$\| X^* - X^{(k)} \|_1 = \sum_{i=1}^{n} | x_i^* - x_i^{(k)} | \leqslant \frac{q^k}{1-q} \sum_{i=1}^{n} | x_i^{(1)} - x_i^{(0)} |$$

相应于∞ - 范数误差估计式为

$$\| X^* - X^{(k)} \|_\infty = \max_i | x_i^* - x_i^{(k)} | \leqslant \frac{q^k}{1-q} \max_i | x_i^{(1)} - x_i^{(0)} |$$

上式虽然给出了误差估计式，但一般来说估计偏大，且计算比较复杂，不实用。当利用电子计算机进行迭代求解时，通常用相邻两次迭代值之差的绝对值来判断是否满足精度要求，决定是否终止迭代过程。即如果

$$\max_i | x_i^{(k)} - x_i^{(k-1)} | \leqslant \varepsilon$$

则取 $X^{(k)}$ 作为方程组解 X^* 的近似值，结束迭代过程。其中 ε 为一个小的正数，代表精度要求。

三、迭代形式的形成

对于给定的方程组 $MX = G$，设法把它化为 $X = AX + B$ 的形式，只要其满足定理条件，即 $\| A \| = q < 1$，就可进行迭代求解。若迭代式

$$X^{(k)} = AX^{(k-1)} + B \quad (k = 1, 2, \cdots)$$

用分量的形式表示，则第 k 次迭代中，$X^{(k)}$ 的第 i 个分量可用下式计算

$$x_i^{(k)} = \sum_{j=1}^{n} a_{ij} x_j^{(k-1)} + b_j \quad (i = 1, 2, \cdots, n; k = 1, 2, \cdots)$$

以上给出的两种具体范数 $\| A \|_\infty$ 和 $\| A \|_1$，只有当矩阵 A 的各元素的绝对值 $| a_{ij} |$ 较小时，才有可能满足 $\| A \| < 1$ 的要求。按此条件，下边针对矩阵 M 的特点给出两种常用的变化方法。

方法 1 如果矩阵 M 中每个对角元素 m_{ii} 的绝对值大于同行上其他元素，那么只要把方程组中第 i 个方程含 x_i 的项留在左边，其他项移到右边，再除以 m_{ii} 即可得

$$X = AX + B$$

其中

$$A = \begin{bmatrix} 0 & -\frac{m_{12}}{m_{11}} & -\frac{m_{13}}{m_{11}} & \cdots & -\frac{m_{1n}}{m_{11}} \\ -\frac{m_{21}}{m_{22}} & 0 & -\frac{m_{23}}{m_{22}} & \cdots & -\frac{m_{2n}}{m_{22}} \\ \vdots & \vdots & \vdots & & \vdots \\ -\frac{m_{n1}}{m_{nn}} & -\frac{m_{n2}}{m_{nn}} & -\frac{m_{n3}}{m_{nn}} & \cdots & 0 \end{bmatrix}, \quad B = \left(\frac{g_1}{m_{11}}, \frac{g_2}{m_{22}}, \cdots\cdots, \frac{g_n}{m_{nn}} \right)^T$$

事实上，它是将 M 分裂为

$$M=\begin{bmatrix} m_{11} & & & \\ & m_{22} & & \\ & & \ddots & \\ & & & m_{nn} \end{bmatrix}-\begin{bmatrix} 0 & -m_{12} & \cdots & -m_{1n} \\ -m_{21} & 0 & \cdots & -m_{2n} \\ \vdots & \vdots & & \vdots \\ -m_{n1} & -m_{n2} & \cdots & 0 \end{bmatrix}=D-L$$

则

$$(D-L)X=G$$

$$DX=LX+G$$

由于 $\det D\neq 0$，所以 D^{-1} 存在，于是

$$X=D^{-1}LX+D^{-1}G$$

即

$$X=AX+B$$

其中，$A=D^{-1}L$，$B=D^{-1}G$。

这样，迭代公式的分量形式为

$$x_i^{(k)}=\frac{g_i}{m_{ii}}-\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{m_{ij}}{m_{ii}}x_i^{(k-1)}\quad (i=1,2,\cdots,n)$$

这种迭代法也称为雅可比迭代法。其迭代格式简单，对于严格对角优势矩阵，满足收敛条件。它的计算步骤为：

(1)给定系数矩阵元素 m_{ij} 及右端自由项 $g_i(i,j=1,2,\cdots,n)$ 和精度要求 ε 或 δ 及初值 $X^{(0)}$；

(2)计算解序列 X

$$x_i=\frac{g_i}{m_{ii}}-\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{m_{ij}}{m_{ii}}x_j^{(0)}\quad (i=1,2,\cdots,n);$$

(3)若 $|x_i-x_i^{(0)}|\leqslant\varepsilon$ 或 $(x_i-x_i^{(0)})/x_i\leqslant\delta$ $(i=1,2,\cdots,n)$，则输出计算结果 $X=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，停止计算；否则令 $x_i^{(0)}=x_i(i=1,2,\cdots,n)$，同第(2)步。

例 4-7 利用迭代法求解方程组

$$\begin{cases} 4x_1+2x_2=2 \\ 2x_1+10x_2+4x_3=6 \\ 4x_2+5x_3=5 \end{cases}$$

要求精度 $\varepsilon=10^{-1}$。

解 该方程组系数矩阵 $M=\begin{bmatrix} 4 & 2 & 0 \\ 2 & 10 & 4 \\ 0 & 4 & 5 \end{bmatrix}$ 中主对角线上的元素的绝对值大于同行上的其他元素。于是，进行雅可比变换得

$$A=\begin{bmatrix} 0 & -0.5 & 0 \\ -0.2 & 0 & -0.4 \\ 0 & -0.8 & 0 \end{bmatrix},\quad B=(0.5,0.6,1.0)^{\mathrm{T}}$$

选取 $X^{(0)}=(0,0,0)$，经 16 次迭代得

$$x_1=0.414,\quad x_2=0.172,\quad x_3=0.862$$

方法 2 若矩阵 M 中，对角线上元素接近于 1，其他元素接近于 0，那么可以作如下转化：

$$X=X-MX+G=(E-M)X+G=AX+B$$

则

$$A=\begin{bmatrix}1-m_{11} & -m_{12} & \cdots & -m_{1n}\\ -m_{21} & 1-m_{22} & \cdots & -m_{2n}\\ \vdots & \vdots & & \vdots\\ -m_{n1} & -m_{n2} & \cdots & 1-m_{nn}\end{bmatrix}$$

迭代公式的分量形式为

$$x_i^{(k)}=(1-m_{ii})x_i^{(k-1)}-\sum_{\substack{j=1\\ j\neq i}}^{n}m_{ij}x_i^{(k-1)}+g_i\quad(i=1,2,\cdots,n)$$

该方法的计算步骤为：

(1)给定 $m_{ij}(i,j=1,2,\cdots,n)$、精度要求 ε 或 δ 以及初值

$$X^{(0)}=(x_1^{(0)},x_2^{(0)},\cdots,x_n^{(0)})^{\mathrm{T}};$$

(2)计算解序列 X

$$x_i=(1-m_{ii})x_i^{(0)}-\sum_{\substack{j=1\\ j\neq i}}^{n}m_{ij}x_j^{(0)}+g_i\quad(i=1,2,\cdots,n);$$

(3)若 $|x_i-x_i^{(0)}|\leqslant\varepsilon$ 或 $|(x_i-x_i^{(0)})/x_i|\leqslant\delta$ $(i=1,2,\cdots,n)$，则输出 $X=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，停止计算；否则，令 $x_i^{(0)}=x_i(i=1,2,\cdots,n)$，回第(2)步。

例 4-8 有一个三组分混合物的水溶液，经分光光度分析得到如表 4-2 中的数据。若给定组分服从比尔规律，并且混合溶液中 3 种吸收物质相互无干扰，求水溶液中各组分的浓度。

表 4-2

λ/nm	组分 A/(L·mol^{-1})	组分 B/(L·mol^{-1})	组分 C/(L·mol^{-1})	总吸收率 D_λ
420	0.03	0.04	1.23	0.087 6
540	1.05	0.06	0.63	0.122 3
600	0.73	1.03	0.08	0.138 0

解 设三组分在水溶液中的浓度分别为 c_A、c_B、c_C，则由比尔定律可得

$$\begin{cases}0.03c_A+0.04c_B+1.23c_C=0.087\,6\\ 1.05c_A+0.06c_B+0.63c_C=0.122\,3\\ 0.73c_A+1.03c_B+0.08c_C=0.138\,0\end{cases}$$

于是,其系数矩阵为

$$M=\begin{bmatrix}0.03 & 0.04 & 1.23\\ 1.05 & 0.06 & 0.63\\ 0.73 & 1.03 & 0.08\end{bmatrix}$$

该矩阵并不符合主对角线上元素的值接近于1,其他元素接近于零的情形。但可将方程组改写为

$$\begin{cases}1.05c_A+0.06c_B+0.63c_C=0.1223\\ 0.73c_A+1.03c_B+0.08c_C=0.1380\\ 0.03c_A+0.04c_B+1.23c_C=0.0876\end{cases}$$

即对矩阵进行行交换,变换时注意右端项要随之一起交换,此时系数矩阵为

$$M=\begin{bmatrix}1.05 & 0.06 & 0.63\\ 0.73 & 1.03 & 0.08\\ 0.03 & 0.04 & 1.23\end{bmatrix}$$

符合上述情形。这样,经交换得

$$A=\begin{bmatrix}-0.05 & -0.06 & -0.63\\ -0.73 & -0.03 & -0.08\\ -0.03 & -0.04 & -0.23\end{bmatrix},\quad B=(0.1223,0.138,0.0876)^{\mathrm{T}}$$

取 $X^{(0)}=(0.1,0.1,0.1)^{\mathrm{T}}$,迭代8次得

$$c_A=0.07188,\quad c_B=0.7782,\quad c_C=0.06694$$

在实际计算过程中,有时所给的方程组 $MX=G$ 初看起来并不满足以上两种情形,如例4-8,但只要改变一下各方程的相对位置,或作适当的初等变换,仍可得到适合上述情况的方程组。

普通迭代法的优点在于:一方面,它具有简单统一的算式 $X=AX+B$,实际计算时,每个方程又是相对独立进行的,所以计算简便;另一方面,由于初始向量选取的任意性,只要矩阵 A 满足收敛条件,计算结果不会受到误差积累的影响,因为任何一次近似值均可视为下一次迭代的初始近似值。

在使用普通迭代法时,初始近似值可选取零项量或右端自由量。

第八节　高斯—赛德尔迭代法

一、方法介绍

高斯—赛德尔迭代法是对普通迭代法的一种改进。它与普通迭代法的区别仅在于:当计算第 k 次迭代各分量的近似值 x_i 时,已经求出新分量 $x_1^{(k-1)},x_2^{(k-1)},\cdots,x_{i-1}^{(k-1)}$。当迭代收敛时,这些新分量要比老分量 $x_1^{(k-1)},x_2^{(k-1)},\cdots,x_{i-1}^{(k-1)}$ 更接近于精确解 $x_1^*,x_2^*,\cdots,$

x_{i-1}^{*}。若用新分量代替对应的老分量进行迭代，则可使迭代过程加速。这样便构成了求解线性方程组 $X=AX+B$ 的高斯—赛德尔迭代格式：

$$x_i^{(k)}=\sum_{j=1}^{i-1}a_{ij}x_j^{(k)}+\sum_{j=i+1}^{n}a_{ij}x_j^{(k-1)}+b_i \quad (i=1,2,\cdots n)$$

将上式展开成方程的一般形式为

$$\begin{cases} x_1^{(k)}=a_{11}x_1^{(k-1)}+a_{12}x_2^{(k-1)}+\cdots+a_{1n}x_n^{(k-1)}+b_1 \\ x_2^{(k)}=a_{21}x_1^{(k)}+a_{22}x_2^{(k-1)}+\cdots+a_{2n}x_n^{(k-1)}+b_2 \\ x_3^{(k)}=a_{31}x_1^{(k)}+a_{32}x_2^{(k)}+\cdots+a_{3n}x_n^{(k-1)}+b_3 \\ \cdots\cdots \\ x_n^{(k)}=a_{n1}x_1^{(k)}+a_{n2}x_2^{(k)}+\cdots+a_{nn}x_n^{(k-1)}+b_n \end{cases}$$

二、高斯—赛德尔迭代法的矩阵形式

若把方程组 $X=AX+B$ 中的矩阵 A 分裂为一个下三角矩阵 L 和一个上三角矩阵 U，即

$$L=\begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_{21} & 0 & \cdots & 0 \\ a_{31} & a_{32} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & 0 \end{bmatrix},\quad U=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

显然

$$A=L+U,\quad X=LX+UX+B$$

于是

$$X^{(k)}=LX^{(k)}+UX^{(k-1)}+B,\quad (E-L)X^{(k)}=UX^{(k-1)}+B$$

因为 $\det(E-L)\neq 0$，故逆矩阵 $(E-L)^{-1}$ 存在。于是

$$X^{(k)}=(E-L)^{-1}UX^{(k-1)}+(E-L)^{-1}B$$

该式说明对矩阵 A 施行高斯—赛德尔迭代法就相当于对矩阵 $(E-L)^{-1}U$ 施行普通迭代法。一般来说，若矩阵 A 是对角优势矩阵，则高斯—赛德尔迭代法和普通迭代法都收敛。当 $\|A\|_{\infty}=\max\limits_{i}\sum\limits_{j=1}^{n}|a_{ij}|<1$ 时，高斯—赛德尔迭代法比普通迭法收敛快。但应指出，高斯—赛德尔迭代法并不是总比普通迭代法好。

例如，方程组 $X=AX+B$。其中，$A=\begin{bmatrix} 2.3 & -5 \\ 1 & -2.3 \end{bmatrix}$，$B=\begin{bmatrix} 3.7 \\ 2.3 \end{bmatrix}$。用普通迭代法收敛，而用高斯—赛德尔迭代法则发散。方程组的精确解为 $X=(1.0,1.0)^{T}$。读者不妨以两种方法试算，分析其结果。

关于高斯—赛德尔迭代法的计算步骤可概括为：

(1)给定 $m_{ij}(i,j=1,2,\cdots,n)$ 和精度 ε 或 δ;

(2)将 $MX=G$ 转化为 $X=AX+B$,并选取初值

$$X^{(0)}=(x_1^{(0)},x_2^{(0)},\cdots,x_n^{(0)})^{\mathrm{T}};$$

(3)计算解序列 X

$$x_i=\sum_{j=1}^{i-1}a_{ij}x_j+\sum_{j=i+1}^{n}a_{ij}x_j^{(0)}+b_i\quad(i=1,2,\cdots,n);$$

(4)若 $|x_i-x^{(0)}|\leqslant\varepsilon$ 或 $|(x_i-x_i^{(0)})/x_i|\leqslant\delta\ (i=1,2,\cdots,n)$,则输出 $X=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$,终止计算;否则,令 $x_i^{(0)}=x_i(i=1,2,\cdots,n)$,回第(3)步。

例4-9 试用高斯—赛德尔迭代法解方程组

$$\begin{bmatrix}3&-5&47&20\\11&16&17&10\\56&22&11&-18\\17&66&-12&7\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}18\\26\\34\\82\end{bmatrix}$$

要求精度 $\varepsilon=10^{-4}$。

解 首先变换方程的次序为

$$\begin{bmatrix}56&22&11&-18\\17&66&-12&7\\3&-5&47&20\\11&16&17&10\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}34\\82\\18\\26\end{bmatrix}$$

现除第4行外,主对角线上的元素均大于其他元素,将其转化为 $X=AX+B$ 的形式得迭代格式为

$$A=\begin{bmatrix}0&-\frac{22}{56}&-\frac{11}{56}&\frac{18}{56}\\-\frac{17}{66}&0&\frac{12}{66}&-\frac{7}{66}\\-\frac{3}{47}&\frac{5}{47}&0&-\frac{20}{47}\\-\frac{11}{10}&-\frac{16}{10}&-\frac{17}{10}&0\end{bmatrix},\quad B=\begin{bmatrix}\frac{34}{56}\\\frac{82}{66}\\\frac{18}{47}\\\frac{26}{10}\end{bmatrix}$$

$$\begin{cases}x_1=(34-22x_2^{(0)}-11x_3^{(0)}+18x_4^{(0)})/56\\x_2=(82-17x_1+12x_3^{(0)}-7x_4^{(0)})/66\\x_3=(18-3x_1+5x_2-20x_4^{(0)})/47\\x_4=(26-11x_1-16x_2-17x_3)/10\end{cases}$$

取初值 $x^{(0)}=(1,1,1,1)^{\mathrm{T}}$ 进行迭代,经35次迭代得到满足精度要求的近似解为

$$x=(-1.076\,89,1.990\,03,1.474\,48,-1.906\,08)^{T}$$

此例中的情况，普通迭代法也可能收敛，读者可采用普通迭代法进行计算，并与上述算法比较迭代的次数。

第九节　松弛迭代法

松弛法简称 SOR 法，是高斯—赛德尔迭代法的加速方法，适宜于求解大型稀疏矩阵方程组。

对于线性方程组 $MX=G$，若 $m_{ii}\neq 0$ $(i=1,2,\cdots,n)$，则可应用高斯—赛德尔迭代法进行计算：

$$x_i^{(k)*}=\frac{1}{m_{ii}}\left[g_i-\sum_{j=1}^{i-1}m_{ij}x_j^{(k)}-\sum_{j=i+1}^{n}m_{ij}x_j^{(k-1)}\right]\quad(i=1,2,\cdots,n;k=1,2,\cdots)$$

现引入松弛因子 ω，利用 $x_i^{(k-1)}$ 和 $x^{(k)*}$ 两个信息相组合来改善第 k 次迭代解 $x_i^{(k)}$，作线性加速：

$$x_i^{(k)}=\omega x_i^{(k)*}+(1-\omega)x_i^{(k-1)}=x_i^{(k-1)}+\omega[x_i^{(k)*}-x_i^{(k-1)}]$$

则

$$x_i^{(k)}=x_i^{(k-1)}+\frac{\omega}{m_{ii}}\left(g_i-\sum_{j=1}^{i-1}m_{ij}x_j^{(k)}-\sum_{j=i}^{n}m_{ij}x_i^{(k-1)}\right)\quad(i=1,2,\cdots,n;k=1,2,\cdots)$$

此式即松弛迭代格式。显然，当 $\omega=1$ 时，此即高斯—赛德尔迭代格式；当 $0<\omega<1$ 时，$x_i^{(k)}$ 为 $x_i^{(k-1)}$ 与 $x_i^{(k)*}$ 的加权平均值，称为低松弛迭代，可改善迭代过程的收敛性；当 $1<\omega<2$ 时 $x_i^{(k)}$ 为 $x_i^{(k-1)}$ 与 $x_i^{(k)*}$ 的外推值，称为超松弛迭代，可加速迭代过程。

若将系数矩阵 M 分解为 $M=D-L-U$，其中

$$D=\begin{bmatrix}m_{11}&&&\\&m_{22}&&\\&&\ddots&\\&&&m_{nn}\end{bmatrix},\quad L=\begin{bmatrix}0&0&\cdots&0\\-m_{21}&0&\cdots&0\\-m_{31}&-m_{32}&\cdots&0\\\vdots&\vdots&&\vdots\\-m_{n1}&-m_{n2}&\cdots&0\end{bmatrix},$$

$$U=\begin{bmatrix}0&-m_{12}&-m_{13}&\cdots&-m_{1n}\\0&0&-m_{23}&\cdots&-m_{2n}\\\vdots&\vdots&\vdots&&\vdots\\0&0&0&\cdots&0\end{bmatrix}$$

现将松弛迭代格式改写为

$$m_{ii}x_i^{(k)}=(1-\omega)m_{ii}x_i^{(k-1)}+\omega\left[g_i-\sum_{j=1}^{i-1}m_{ij}x_j^{(k)}-\sum_{j=i+1}^{n}m_{ij}x_j^{(k-1)}\right]$$

于是

$$DX^{(k)}=(1-\omega)DX^{(k-1)}+\omega[G+LX^{(k)}+UX^{(k-1)}]$$

$$(D-\omega L)X^{(k)}=[(1-\omega)D+\omega U]X^{(k-1)}+\omega G$$

由于 $\det(D-\omega L)\neq 0$，故 $(D-\omega L)^{-1}$ 存在，则

$$X^{(k)}=(D-\omega L)^{-1}[(1-\omega)D+\omega U]X^{(k-1)}+\omega(D-\omega L)^{-1}G$$

这说明，对矩阵 A 施行松弛迭代就相当于对矩阵

$$(D-\omega L)^{-1}[(1-\omega)D+\omega U]$$

施行普通迭代。这样，松弛迭代法的矩阵形式可写为

$$\begin{cases}X=AX+B\\ A=(D-\omega L)^{-1}[(1-\omega)D+\omega U]\\ B=\omega(D-\omega L)^{-1}G\end{cases}$$

关于松弛因子 ω 的选择是较困难的，一般可通过计算机计算确定。

综上所述，松弛迭代法的计算步骤可概括为：

(1)给定系数矩阵元素 $m_{ij}(i,j=1,2,\cdots,n)$、计算精度 ε 或 δ 和初值

$$X^{(0)}=(x_1^{(0)},X_2^{(0)},\cdots,x_n^{(0)})^{\mathrm{T}};$$

(2)计算解序列 X(先设定一个松弛因子 ω，再在计算过程中调整)

$$x_i=x_i^{(0)}+\frac{\omega}{m_{ii}}\left(g_i-\sum_{j=1}^{i-1}m_{ij}x_j-\sum_{j=i}^{n}m_{ij}x_j^{(0)}\right)\quad(i=1,2,\cdots,n);$$

(3)若 $|x_i-x_i^{(0)}|\leqslant\omega$ 或 $|(x_i-x_i^{(0)})/x_i|\leqslant\delta$ $(i=1,2,\cdots,n)$，则输出计算结果 $X=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，终止计算；否则，令 $x_i^{(0)}=x_i(i=1,2,\cdots,n)$，回第(2)步。

例 4-10 利用松弛迭代法解方程组

$$\begin{bmatrix}8&-3&2\\4&11&-1\\6&3&12\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}20\\33\\36\end{bmatrix}$$

要求精度 $\varepsilon=10^{-4}$。

解 根据给定的系数矩阵，其松弛迭代格式为

$$x_1=x_1^{(0)}+\frac{\omega}{8}(20-8x_1^{(0)}+3x_2^{(0)}-2x_3^{(0)})$$

$$x_2=x_2^{(0)}+\frac{\omega}{11}(33-4x_1-11x_2^{(0)}+x_3^{(0)})$$

$$x_3=x_3^{(0)}+\frac{\omega}{12}(36-x_1-3x_2-12x_3^{(0)}$$

取初值 $x^{(0)}=(1,1,1)$，ω 从 0.50 开始，每次增量为 0.05，直到 1.10，得出 $\omega=0.95$ 时，经 5 次迭代可得

$$x_1=3.000\,0,\quad x_2=2.000\,0,\quad x_3=1.000\,0$$

习　题

1. 用高斯消去法解下列方程组：

(1) $\begin{cases} 2.51x_1 + 1.48x_2 + 4.53x_3 = 0.05 \\ 1.48x_1 + 0.93x_2 - 1.30x_3 = 1.03 \\ 2.68x_1 + 3.04x_2 - 1.48x_3 = -0.53 \end{cases}$

(2) $\begin{cases} 2x_1 + 2x_2 + 4x_3 - 2x_4 = 10 \\ x_1 + 3x_2 + 2x_3 + x_4 = 17 \\ 3x_1 + x_2 + 3x_3 + x_4 = 18 \\ x_1 + 3x_2 + 4x_3 + 2x_4 = 27 \end{cases}$

2. 用高斯主元素消去法解下列方程组：

(1) $\begin{cases} x_1 + 2x_2 - 12x_3 + 8x_4 = 27 \\ 5x_1 + 4x_2 + 7x_3 - 2x_4 = 4 \\ -3x_1 + 7x_2 + 9x_3 + 5x_4 = 11 \\ 6x_1 - 12x_2 - 9x_3 + 3x_4 = 48 \end{cases}$

(2) $\begin{cases} 5x_1 + 4x_2 + 2x_3 = 7 \\ 2x_1 + x_2 + 4x_3 = 3 \\ 4x_1 + 3x_2 + 6x_3 = 2 \end{cases}$

3. 用三次样条插值求某化学反应生成物的瞬态浓度时，得到如下对角方程组：

$$\begin{cases} 3c_1 + c_2 = -3.6 \\ c_1 + 4c_2 + c_3 = -2.25 \\ 2c_2 + 10c_3 = -1.1 \end{cases}$$

试用追赶法求解。

4. 用追赶法求解方程组：

$$\begin{cases} 2x_1 - x_2 = 1 \\ -x_1 + 2x_2 - x_3 = 0 \\ -x_2 + 2x_3 - x_4 = 0 \\ -x_3 + 2x_4 - x_5 = 0 \\ -x_4 + 2x_5 = 7 \end{cases}$$

5. 利用 LU 分解法解下列方程组：

(1) $\begin{cases} 2x_1 + 4x_2 - 2x_3 = 6 \\ x_1 - x_2 + 5x_3 = 0 \\ 4x_1 + x_2 - 2x_3 = 2 \end{cases}$

(2) $\begin{cases} 3x_1 + 2x_2 - x_3 = 10 \\ x_1 - x_2 + 3x_3 = -4 \\ 2x_1 + x_2 - 3x_3 = 16 \end{cases}$

6. 利用 LDL^{T} 分解法解下列方程组：

(1) $\begin{cases} 5x_1 - x_2 = 9 \\ -x_1 + 5x_2 - x_3 = 4 \\ -x_2 + 5x_3 = -6 \end{cases}$

(2) $\begin{cases} -6x_1 + x_2 + x_3 = -12 \\ x_1 - 6x_2 + x_3 = -32 \\ x_1 + x_2 - 6x_3 = -42 \end{cases}$

7. 利用平方根法解方程组：

$$\begin{cases} 5x_1 + 7x_2 + 6x_3 + 5x_4 = 23 \\ 7x_1 + 10x_2 + 8x_3 + 7x_4 = 32 \\ 6x_1 + 8x_2 + 10x_3 + 9x_4 = 33 \\ 5x_1 + 7x_2 + 9x_3 + 10x_4 = 31 \end{cases}$$

8. 利用普通迭代法求解下列方程组($\varepsilon = 10^{-4}$)：

(1) $\begin{cases} 10x_1 - x_2 - 2x_3 = 7.2 \\ -x_1 + 10x_2 - 2x_3 = 8.3 \\ -x_1 - x_2 + 5x_3 = 4.2 \end{cases}$

(2) $\begin{cases} x_1 + 0.02x_2 + 0.05x_3 = 0.97 \\ 0.05x_1 + x_2 + 0.037\,5x_3 = 0.987\,5 \\ 0.012\,5x_1 + 0.075x_2 + 0.937\,5x_3 = 0.85 \end{cases}$

9. 利用高斯—赛德尔迭代法解方程组：

$$\begin{cases} 9x_1 - 5x_3 = 10 \\ 2x_2 - 12x_3 = -2 \\ -5x_1 - 12x_2 + 20x_3 = 0 \end{cases}$$

要求精度 $\varepsilon = 10^{-3}$。

10. 现要求精度 $\varepsilon = 10^{-4}$，试分别用普通迭代法、高斯—赛德尔迭代法和松弛迭代法解方程组：

$$\begin{bmatrix} -2 & 1 & & & & & & & & \\ 1 & -2 & 1 & & & & & & & \\ & 1 & -2 & 1 & & & & & & \\ & & 1 & -2 & 1 & & & & & \\ & & & 1 & -2 & 1 & & & & \\ & & & & 1 & -2 & 1 & & & \\ & & & & & 1 & -2 & 1 & & \\ & & & & & & 1 & -2 & 1 & \\ & & & & & & & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \end{bmatrix} = \begin{bmatrix} -0.5 \\ -1.5 \\ -1.5 \\ -1.5 \\ -1.5 \\ -1.5 \\ -1.5 \\ -1.5 \\ -1.5 \\ -0.5 \end{bmatrix}$$

并比较其收敛速度。

第五章　函数的多项式插值

在实际工程计算中，经常会遇到许多以表格形式给出的函数。例如，化工物性数据表、三角函数表、对数表、方根表、特殊函数表等，它们都是将自变量与函数的关系通过表格的形式给出。这些表格函数没有直接给出未列点处的函数值，也不便于微分和积分的计算。插值问题就是要通过表格函数中若干点数据构造一个比较简单的函数，来近似表达原来的函数，从而得到函数的某些近似值。在这方面成熟的方法很多，但最常用的是代数多项式，因为它形式简单，也便于微分和积分。本章仅讨论代数多项式插值法，简称代数插值法或多项式插值法。

第一节　概　述

代数插值法可描述为：给定函数 $y=f(x)$ 在区间 $[a,b]$ 上的 $n+1$ 个点 $a\leqslant x_0<x_1<x_2<\cdots<x_n\leqslant b$ 及其上的函数值 $y_i=f(x_i)(i=0,1,2,\cdots,n)$，建立一个次数不超过 n 的代数多项式

$$P_n(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n$$

使其满足

$$P_n(x_i)=f(x_i)\quad(i=0,1,2,\cdots,n)$$

其中，a_i 为实数，则称 $P_n(x)$ 为函数 $f(x)$ 的插值多项式，称 $x_0,x_1,x_2,\cdots,x_n$ 为其插值节点，$[a,b]$ 为其插值区间，称离散的表格函数 $y=f(x)$ 为被插值函数。

插值法的几何意义为：通过给定的 $n+1$ 个几何节点 $(x_i,y_i)(i=0,1,2,\cdots,n)$，作一条 n 次多项式曲线 $y=P_n(x)$ 来近似地代替曲线 $y=f(x)$，如图 5-1 所示。显然有

$$\begin{cases}f(x_i)=P_n(x_i) & (i=0,1,2,\cdots,n)\\ R_n(x)=f(x)-P_n(x) & (\text{非节点处})\end{cases}$$

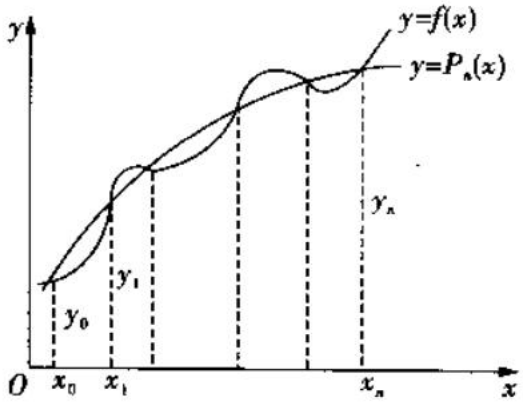


图 5-1

即在非节点处 $P_n(x) \approx f(x)$，它们之间存在有偏差 $R_n(x)$，称 $R_n(x)$ 为插值多项式的余项。

容易证明插值问题的解是唯一的。用反证法，假如有两个 n 次多项式 $y=P_n(x)$ 与 $y=q_n(x)$ 均满足

$$f(x_i)=P_n(x_i)=q_n(x_i)$$

令

$$F_n(x)=P_n(x)-q_n(x)$$

有

$$F_n(x_i)=0 \quad (i=0,1,2,\cdots,n)$$

即不超过 n 次的多项式 $F_n(x)$ 有 $n+1$ 个零点，由此可断定 $F_n(x)=0$，即 $P_n(x)=q_n(x)$，所以满足插值条件的插值多项式是唯一的。

第二节　拉格朗日插值多项式

拉格朗日插值多项式是一族插值的基本公式，即 n 次插值多项式，是代数插值最基本而常用的一类公式。

一、线性插值多项式

线性插值多项式也称为直线插值公式，即一次插值公式

$$n=1, \quad y=P_1(x)=a_0+a_1x$$

实际上，线性插值多项式就是构造一个直线函数 $y \approx P_1(x)$ 来近似表示被插值函数 $y=f(x)$。按照插值条件有

$$y_0=P_1(x_0)=a_0+a_1x_0$$
$$y_1=P_1(x_1)=a_0+a_1x_1$$

由以上两式得

$$a_0=(y_0x_1-y_1x_0)/(x_1-x_0), \quad a_1=(y_1-y_0)/(x_1-x_0)$$

所以

$$y=P_1(x)=\frac{y_0x_1-y_1x_0}{x_1-x_0}+\frac{y_1-y_0}{x_1-x_0}x$$

现将该式整理为

$$y=P_1(x)=y_0\frac{x-x_1}{x_0-x_1}+y_1\frac{x-x_0}{x_1-x_0}$$

令

$$L_0(x)=(x-x_1)/(x_0-x_1), \quad L_1(x)=(x-x_0)/(x_1-x_0)$$

于是

$$P_1(x) = y_0L_0(x) + y_1L_1(x) = \sum_{k=0}^{1} y_kL_k(x)$$

显然，$L_0(x)$ 和 $L_1(x)$ 都是关于 x 的线性函数，称为线性插值的基函数，线性插值多项式 $P_1(x)$ 就是这两个线性插值基函数的线性组合。因此，只要知道基函数 $L_k(x)(k=0,1)$ 的基本特征，便可知道插值函数 $y=P_1(x)$ 的基本特征。

线性插值函数在节点 x_0 和 x_1 处的基本特征为：

$$L_k(x)=\frac{x-x_j}{x_k-x_j}=\begin{cases}1, & x=x_k\\ 0, & x=x_j\end{cases} \quad (k,j=0,1,\text{且 } k\neq j)$$

因此，对线性插值多项式 $P_1(x) = \sum_{k=0}^{1} y_kL_k(x)(k=0,1)$，有

$$\begin{cases}P_1(x)=f(x) & (x=x_k)\\ P_1(x)\approx f(x) & (x\neq x_k)\end{cases}$$

线性插值多项式主要用于求取表格中未列点处的表格函数值。

例 5-1 已知 CH_4 在 400 K 和 500 K 下的标准焓变分别为 $\Delta H^0_{400}=3.887\ 4$ kJ/mol 和 $\Delta H^0_{500}=8.234\ 6$ kJ/mol，试求其在 453 K 下的标准焓变。

解 由线性插值公式可得

$$\Delta H^0_{453}=3.887\ 4\frac{453-500}{400-500}+8.234\ 6\frac{453-400}{500-400}=6.191\ 4\ (\text{kJ/mol})$$

二、二次插值多项式

二次插值多项式也称为抛物插值公式，即

$$n=2,\quad y=P_2(x)=a_0+a_1x+a_2x^2$$

此时需要 3 个节点 (x_0,y_0)，(x_1,y_1)，(x_2,y_2)，按照插值条件有

$$y_0=P_2(x_0)=a_0+a_1x_0+a_2x_0^2$$
$$y_1=P_2(x_1)=a_0+a_1x_1+a_2x_1^2$$
$$y_2=P_2(x_2)=a_0+a_1x_2+a_2x_2^2$$

由以上三式得

$$a_0=\frac{\begin{vmatrix}y_0 & x_0 & x_0^2\\ y_1 & x_1 & x_1^2\\ y_2 & x_2 & x_2^2\end{vmatrix}}{\begin{vmatrix}1 & x_0 & x_0^2\\ 1 & x_1 & x_1^2\\ 1 & x_2 & x_2^2\end{vmatrix}},\quad a_1=\frac{\begin{vmatrix}1 & y_0 & x_0^2\\ 1 & y_1 & x_1^2\\ 1 & y_2 & x_2^2\end{vmatrix}}{\begin{vmatrix}1 & x_0 & x_0^2\\ 1 & x_1 & x_1^2\\ 1 & x_2 & x_2^2\end{vmatrix}},\quad a_2=\frac{\begin{vmatrix}1 & x_0 & y_0\\ 1 & x_1 & y_1\\ 1 & x_2 & y_2\end{vmatrix}}{\begin{vmatrix}1 & x_0 & x_0^2\\ 1 & x_1 & x_1^2\\ 1 & x_2 & x_2^2\end{vmatrix}}$$

整理得

$$y = P_2(x) = y_0 \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} + y_1 \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} + y_2 \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}$$

即

$$y = P_2(x) = \sum_{k=0}^{1} y_k L_k(x)$$

其中，二次插值基函数为

$$L_k(x) = \prod_{\substack{j=0 \\ j \neq k}}^{2} \frac{x - x_j}{x_k - x_j} \quad (k = 0,1,2)$$

其特征为

$$L_k(x) = \prod_{\substack{j=0 \\ j \neq k}}^{2} \frac{x - x_j}{x_k - x_j} = \begin{cases} 1, & x = x_k \\ 0, & x = x_j \end{cases} \quad (k = 0,1,2)$$

因此，二次插值多项式

$$P_2(x) = \sum_{k=0}^{2} y_k L_k(x) = \begin{cases} f(x_k), & x = x_k \\ f(x), & x \neq x_k \end{cases} \quad (k = 0,1,2)$$

二次插值多项式也常用于求取表格中未列点处的表格函数值，其精度高于线性插值公式。

例 5-2 用二次插值计算 sin0.336 7 的值。已知 sin0.32 = 0.314 567，sin0.34 = 0.333 487，sin0.36 = 0.352 274。

解 由抛物插值多项式得

$$\begin{aligned} \sin 0.336\,7 = {} & 0.314\,567 \frac{(0.336\,7-0.34)(0.336\,7-0.36)}{(0.32-0.34)(0.32-0.36)} + \\ & 0.333\,487 \frac{(0.336\,7-0.32)(0.336\,7-0.36)}{(0.34-0.32)(0.34-0.36)} + \\ & 0.352\,274 \frac{(0.336\,7-0.32)(0.336\,7-0.34)}{(0.36-0.32)(0.36-0.34)} \\ = {} & 0.330\,374 \end{aligned}$$

该结果与 6 位正弦表查得的数据完全一致。说明抛物插值的精度是相当高的。读者可用线性插值计算进行比较。

三、拉格朗日插值多项式

1. 公式导出

上面讨论了线性插值多项式和二次插值多项式的构造。

当 $n = 1$ 时，需要 2 个节点 (x_0, y_0)，(x_1, y_1)，得

$$y = P_1(x) = \sum_{k=0}^{1} y_k L_k(x)$$

其中

$$L_k(x)=\frac{x-x_j}{x_k-x_j}\quad(k,j=0,1;k\neq j)$$

当 $n=2$ 时，需要 3 个节点 $(x_0,y_0),(x_1,y_1),(x_2,y_2)$，得

$$y=P_2(x)=\sum_{k=0}^{2}y_kL_k(x)$$

其中

$$L_k(x)=\prod_{\substack{j=0\\j\neq k}}^{2}\frac{x-x_j}{x_k-x_j}\quad(k,j=0,1,2)$$

显然 $L_k(x)$ 具有如下特征：

(1) $L_k(x)=\begin{cases}1, & x=x_k\\0, & x=x_j\end{cases}$ $(k,j=0,1,2$ 且 $k\neq j)$。

(2) 它只与插值节点有关，而与节点处的函数值无关。

(3) 具有对称性，即分子包含除 $(x-x_k)$ 之外的其他所有因子，分母则只把分子中的 x 换成 x_k 即可得到。由于各节点互不相同，所以分母不会为零。

于是，对于一般的 $n+1$ 个节点 $(x_0,y_0),(x_1,y_1),\cdots,(x_n,y_n)$，可构造一个 n 次插值多项式

$$y=P_n(x)=\sum_{k=0}^{2}y_kL_k(x)$$

其中

$$L_k(x)=\prod_{\substack{j=0\\j\neq k}}^{n}\frac{x-x_j}{x_k-x_j}=\begin{cases}f(x_k), & x=x_k\\f(x), & x\neq x_k\end{cases}\quad(k,j=0,1,2,\cdots n)$$

该式称为拉格朗日插值多项式，它由 $n+1$ 个 n 次插值基函数线性组合而成。

2. 拉格朗日插值多项式中的余项

若在区间 $[a,b]$ 上用 $P_n(x)$ 近似代替 $f(x)$，则其截断误差为

$$R_n(x)=f(x)-P_n(x)$$

即截断误差为插值多项式的余项。关于插值余项估计有如下定理。

定理 设函数 $y=f(x)$ 的 n 阶导数 $f^{(n)}(x)$ 在区间 $[a,b]$ 上连续，且它的 $n+1$ 阶导数 $f^{(n+1)}(x)$ 在 (a,b) 内存在，$P_n(x)$ 为 $f(x)$ 在节点 $x_0,x_1,\cdots,x_n$ 处的插值函数，则对于任意的 $x\in[a,b]$，插值多项式的余项

$$R_n(x)=f(x)-P_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{j=0}^{n}(x-x_j)$$

其中，$\xi\in(a,b)$ 且依赖于 x。

证明 设 $f^{(n)}(x)$ 在 $[a,b]$ 上连续，$f^{(n+1)}(x)$ 在 (a,b) 内存在。由

$$R_n(x_k)=f(x_k)-P_n(x_k)=0$$

知 $x_0,x_1,\cdots,x_n$ 都是 $R_n(x)$ 的零点。故可设

$$R_n(x) = K(x)(x-x_0)(x-x_1)\cdots(x-x_n) = K(x)\prod_{k=0}^{n}(x-x_k)$$

其中，$K(x)$为待定函数。

若令

$$\omega(x) = \prod_{k=0}^{n}(x-x_k)$$

则

$$R_n(x) = f(x) - P_n(x) = K(x)\omega(x)$$

为了求得 $K(x)$，作辅助函数

$$F(t) = f(t) - P_n(t) - K(x)\omega(t)$$

显然 $F^{(n)}(t)$在$[a,b]$上连续，$F^{(n+1)}(t)$在(a,b)内存在，且

$$F^{(n+1)}(t) = f^{(n+1)}(t) - K(x)(n+1)!$$

当 $t=x_i(i=0,1,\cdots,n)$时，$F(t)=0$，而对固定的 x，总能选取适当的 $K(x)$使 $F(x)=0$。于是，对于这样选择的 $K(x)$值，$F(t)$在$[a,b]$上至少有 $n+2$ 个零点 $x_0,x_1,\cdots,x_n,x$。由罗尔定理知，$F'(t)$在 $F(t)$的两个零点之间至少有 1 个零点，故 $F'(t)$在$[a,b]$内至少有 $n+1$个互异的零点，对 $F'(t)$再应用罗尔定理，知 $F''(t)$在$[a,b]$上至少有 n 个互异的零点。依此类推，可知 $F^{(n+1)}(t)$在$[a,b]$内至少有一个零点，记为 ξ，于是得

$$F^{(n+1)}(\xi) = f^{(n+1)}(\xi) - K(x)(n+1)! = 0$$

则

$$K(x) = \frac{1}{(n+1)!}f^{(n+1)}(\xi)$$

故

$$R_n(x) = \frac{1}{(n+1)!}f^{(n+1)}(\xi)\omega(x) = \frac{1}{(n+1)!}f^{(n+1)}(\xi)\prod_{k=0}^{n}(x-x_k)$$

因此，只要 $f^{(n)}(x)$在$[a,b]$上连续，$f^{(n+1)}(x)$在(a,b)内存在，且 $P_n(x)$为满足插值条件 $P_n(x_k)=f(x_k)(k=0,1,2,\cdots,n)$的 n 次插值多项式，则对于任何 $x\in[a,b]$，插值余项

$$R_n(x) = f(x) - P_n(x) = \frac{1}{(n+1)!}f^{(n+1)}(\xi)\omega(x)$$

其中，$\xi\in[a,b]$且与 x 有关。

于是，由 $P_n(x)$代替 $f(x)$的误差限为

$$|R_{(n)}(x)| = \max_{a\le x\le b}|f^{(n+1)}(x)|\frac{|\omega(x)|}{(n+1)!}$$

可见，误差限 $|R_{(n)}(x)|$除与 $|f^{(n+1)}(x)|$有关外，还与 $\omega(x)$有关，而 $\omega(x)$与节点的选择密切相关。因此，在进行插值计算时，应尽量选择与被插值点 x 最靠近的 $n+1$ 个节点。

拉格朗日插值多项式的计算很有规律，便于进行计算机编程计算，但它也有一个很大的缺点，即计算工作量比较大，当增加插值节点时，计算插值多项式的工作必须从头做

起。

例 5-3 已查得 CO_2 在水中不同温度下的溶解度为：

T/℃	0	1	3	5
S	0.334 6	0.321 3	0.297 8	0.277 4

试求 CO_2 在 4 ℃时的溶解度。

解 由拉格朗日插值多项式得

$$S_{4℃}=0.334\,6\frac{(4-1)(4-3)(4-5)}{(0-1)(0-3)(0-5)}+0.321\,3\frac{(4-0)(4-3)(4-5)}{(1-0)(1-3)(1-5)}+$$
$$0.297\,8\frac{(4-0)(4-1)(4-5)}{(3-0)(3-1)(3-5)}+0.277\,4\frac{(4-0)(4-1)(4-3)}{(5-0)(5-1)(5-3)}$$
$$=0.287\,3$$

第三节 差分、差商与牛顿插值公式

一、有限差

所谓的有限差是指函数的有限差，包括差分与差商（亦称均差）两个概念。这些概念与数学分析中的微分与微商的概念是平行的，只不过微分与微商适用于连续函数的情形，而差分与差商是对离散函数来说的。电子计算机所处理的数据本质上是离散化的，所以在数值计算的许多领域中差分与差商均起重要的作用。换句话说，它们是把连续函数与离散函数相互转化的重要工具。

（一）差分

1. 差分的定义与差分表

设函数 $y=f(x)$ 在自变量 x 的等距节点 $x_k=x_0+kh(h>0,k=0,1,2,\cdots)$ 的函数值为已知，即

$$y_0=f(x_0),\quad y_1=f(x_1),\quad \cdots$$

则数

$$\Delta y_0=y_1-y_0,\quad \Delta y_1=y_2-y_1,\quad \cdots,\quad \Delta y_l=y_{l+1}-y_l,\quad \cdots$$

称为函数 $f(x)$ 的一阶有限向前差分，简称差分。

用同样的方法可定义二阶差分以及任意阶差分：

$$\Delta^2 y_0=\Delta y_1-\Delta y_0,\quad \Delta^2 y_1=\Delta y_2-\Delta y_1,\quad \cdots,\quad \Delta^2 y_l=\Delta y_{l+1}-\Delta y_l,\quad \cdots,$$
$$\cdots\cdots$$

$$\Delta^k y_0 = \Delta^{k-1} y_1 - \Delta^{k-1} y_0, \quad \Delta^k y_1 = \Delta^{k-1} y_2 - \Delta^{k-1} y_1, \quad \cdots,$$
$$\Delta^k y_l = \Delta^{k-1} y_{l+1} - \Delta^{k-1} y_l, \quad \cdots$$

按差分的定义，函数$f(x)$的各阶差分可列入一个表中，这样的表称为差分表。差分写在计算它时所用的两数之间，如表5－1所示。

表5－1

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$	$\Delta^5 y$
x_0	y_0					
		Δy_0				
x_1	y_1		$\Delta^2 y_0$			
		Δy_1		$\Delta^3 y_0$		
x_2	y_2		$\Delta^2 y_1$		$\Delta^4 y_0$	
		Δy_2		$\Delta^3 y_1$		$\Delta^5 y_0$
x_3	y_3		$\Delta^2 y_2$		$\Delta^4 y_1$	
		Δy_3		$\Delta^3 y_2$		
x_4	y_4		$\Delta^2 y_3$			
		Δy_4				
x_5	y_5					

也可以把差分表压缩成更加紧凑的形式，如表5－2所示。

表5－2

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$	$\Delta^5 y$
x_0	y_0	Δy_0	$\Delta^2 y_0$	$\Delta^3 y_0$	$\Delta^4 y_0$	$\Delta^5 y_0$
x_1	y_1	Δy_1	$\Delta^2 y_1$	$\Delta^3 y_1$	$\Delta^4 y_1$	
x_2	y_2	Δy_2	$\Delta^2 y_2$	$\Delta^3 y_2$		
x_3	y_3	Δy_3	$\Delta^2 y_3$			
x_4	y_4	Δy_4				
x_5	y_5					

2. 差分的重要性质

(1)两个函数和的差分等于各自差分的和。即设$F(x)=f(x)+g(x)$，则

$$\Delta F(x_i) = \Delta f(x_i) + \Delta g(x_i)$$

事实上，设x_i为任意节点，则

$$\begin{aligned}\Delta F(x_i) &= F(x_{i+1}) - F(x_i)\\ &= f(x_{i+1}) + g(x_{i+1}) - f(x_i) - g(x_i)\\ &= f(x_{i+1}) - f(x_i) + g(x_{i+1}) - f(x_i)\\ &= \Delta f(x_i) + \Delta g(x_i)\end{aligned}$$

(2)常数与函数积的差分等于常数乘以函数的差分。即若 $F(x)=Af(x)$,则

$$\Delta F(x)=A\Delta f(x)$$

证明如下:

由于

$$\Delta F(x)=F(x_{i+1})-F(x_i)=Af(x_{i+1})-Af(x_i)$$

所以

$$\Delta F(x)=A\Delta f(x_i)$$

(3)n 次多项式的第 n 阶有限差分等于常数,从而更高阶的有限差分等于零。

现设 k 次多项式为 $P(x)=x^k$,先证明其一阶差分为 $(k-1)$ 次多项式。事实上

$$\Delta P(x_i)=P(x_{i+1})-P(x_i)=x_{i+1}^k-x_i^k=(x_i+h)^k-x_i^k$$

$$=khx_i^{k-1}+\frac{k(k-1)}{2!}h^2x_i^{k-1}+\cdots+h^k$$

再利用性质(1)和性质(2)即可证明。

(4)函数的第 k 阶差分可以由 $k+1$ 个点上的函数值线性表示。即函数在一点上的 k 阶差分可表示为

$$\Delta^k y_1=y_{1+k}-\frac{k}{1!}y_{1+k-1}+\frac{k(k-1)}{2!}y_{1+k-2}+\cdots+(-1)^k y_1$$

这个性质可用数学归纳法来进行证明。

显然,当 $k=1$ 时

$$\Delta y_1=y_2-y_1=y_{1+1}-y_1$$

假定当 $k=k-1$ 时

$$\Delta^{k-1}y_1=y_{1+k-1}-\frac{k-1}{1!}y_{1+k-2}+\frac{(k-1)(k-2)}{2!}y_{1+k-3}+\cdots+(-1)^{k-1}y_1$$

当 $k=k$ 时

$$\Delta^k y_1=\Delta^{k-1}y_2-\Delta^{k-1}y_1$$

所以

$$\Delta^k y_1=y_{2+k-1}-\frac{k-1}{1!}y_{2+k-2}+\frac{(k-1)(k-2)}{2!}y_{2+k-3}+\cdots+(-1)^{k-1}y_2-$$

$$\left[y_{1+k-1}-\frac{k-1}{1!}y_{1+k-2}+\frac{(k-1)(k-2)}{2!}y_{1+k-3}+\cdots+(-1)^{k-1}y_1\right]$$

$$=y_{1+k}-\frac{k}{1!}y_{1+k-1}+\frac{k(k-1)}{2!}y_{1+k-2}+\cdots+(-1)^k y_1$$

于是性质得证。

可见,线性组合的各项系数是正负相间的二项式系数。

(5)各节点上的函数值亦可表示成各阶差分的线性组合。即

$$y_{1+k}=y_1+\frac{k}{1!}\Delta y_1+\frac{k(k-1)}{2!}\Delta^2 y_1+\cdots+\Delta^k y_1$$

读者可以 $k=3$ 的情形验证等式。

3. 函数的误差对各阶差分的影响

如前所述，各阶差分是由一个函数表构造的，通常表中的函数值是带有一定的误差的，不管这种误差的来源如何，都会扩散到各阶差分中去。现仅就一个点 y_t 上带有的误差 ε 进行观察，参见表 5-3。

表 5-3

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
x_{t-3}	y_{t-3}				
		Δy_{t-3}			
x_{t-2}	y_{t-2}		$\Delta^2 y_{t-3}$		$\Delta^4 y_{t-4}+\varepsilon$
		Δy_{t-2}		$\Delta^3 y_{t-3}+\varepsilon$	
x_{t-1}	y_{t-1}		$\Delta^2 y_{t-2}+\varepsilon$		$\Delta^4 y_{t-3}-4\varepsilon$
		$\Delta y_{t-1}+\varepsilon$		$\Delta^3 y_{t-2}+3\varepsilon$	
x_t	$y_t+\varepsilon$		$\Delta^2 y_{t-1}-2\varepsilon$		$\Delta^4 y_{t-2}+6\varepsilon$
		$\Delta y_t-\varepsilon$		$\Delta^3 y_{t-1}+3\varepsilon$	
x_{t+1}	y_{t+1}		$\Delta^2 y_t+\varepsilon$		$\Delta^4 y_{t-1}-4\varepsilon$
		Δy_{t+1}		$\Delta^3 y_t+\varepsilon$	
x_{t+2}	y_{t+2}		$\Delta^2 y_{t+1}$		$\Delta^4 y_t+\varepsilon$
		Δy_{t+2}			
x_{t+3}	y_{t+3}				

由表 5-3 可见，在一个点上的函数的误差，可以影响第 k 阶差分的 $k+1$ 个值。在这 $k+1$ 个差分值中，原始误差 ε 的系数是正负交错的二项式系数 C_k^m

$$C_k^m=\frac{k!}{m!\ (k-m)!}\qquad (\text{这里 } k=4;m=0,1,2,3,4)$$

可见，高阶差分往往会产生很大的误差。因此，在近似计算中应尽量避免使用高阶差分。

4. 函数的向后差分

依照向前差分，函数 $y=f(x)$ 的向后差分可以定义为：

$$\nabla y_t=y_t-y_{t-1},\quad \nabla^k y_t=\nabla^{k-1}y_t-\nabla^{k-1}y_{t-1}\quad (k=2,3,\cdots)$$

用数学归纳法可以证明函数 $y=f(x)$ 的向前与向后差分有如下关系：

(1) $\nabla^k y_t=\Delta^k y_{t-k}$；　　(2) $\Delta^k y_t=\nabla^k y_{t+k}$。

这里仅对关系(2)给予数学归纳法的简略证明。事实上，

$$\Delta y_t=y_{t+1}-y_t=\nabla y_{t+1}$$

设

$$\Delta^{k-1}y_t=\nabla^{k-1}y_{t+k-1}$$

则

$$\Delta^k y_t=\Delta^{k-1}y_{t+1}-\Delta^{k-1}y_t=\nabla^{k-1}y_{t+k}-\nabla^{k-1}y_{t+k-1}=\nabla^k y_{t+k}$$

5. 函数的向中心差分

中心差分是另一种较常用的差分。其定义及算符为：

$$\delta y_t = y_{t+\frac{1}{2}} - y_{t-\frac{1}{2}}$$

对于中心差分 δy_t，用到 $y_{t+\frac{1}{2}}$ 和 $y_{t-\frac{1}{2}}$ 这两个值，它们不是函数表上的值。若用给定的函数表上的值，一阶中心差分应写为

$$\delta y_{t+\frac{1}{2}} = y_{t+1} - y_t, \quad \delta y_{t-\frac{1}{2}} = y_t - y_{t-1}$$

二阶中心差分为

$$\delta^2 y_t = \delta y_{t+\frac{1}{2}} - \delta y_{t-\frac{1}{2}} = y_{t+1} - 2y_t + y_{t-1} = \Delta^2 y_{t-1}$$

（二）差商

对于等距节点的列表函数给出了差分的概念，它将是用于研究函数的特性的有力工具。但是，往往有些列表函数是按不等距节点产生的，对于这类函数差分就不适用了，所以再引入差商及其性质以适用于研究这类函数的需要。

1. 差商的定义

设已知不等距节点 $x_0, x_1, x_2, \cdots, x_k, \cdots$ 的函数值 $y_0, y_1, y_2, \cdots, y_k, \cdots$ 为已知，则数

$$f(x_0, x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

$$f(x_1, x_2) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

$$f(x_2, x_3) = \frac{f(x_3) - f(x_2)}{x_3 - x_2}$$

称为函数 $y = f(x)$ 在所给两点上的一阶差商，一阶差商有明确的几何意义，它是图形 $y = f(x)$ 上过两点的弦与 x 轴夹角的正切。

二阶差商可以用一阶差商来定义，记作

$$f(x_0, x_1, x_2) = \frac{f(x_1, x_2) - f(x_0, x_1)}{x_2 - x_0}$$

三阶差商为

$$f(x_0, x_1, x_2, x_3) = \frac{f(x_1, x_2, x_3) - f(x_0, x_1, x_2)}{x_3 - x_0}$$

一般地，k 阶差商为

$$f(x_0, x_1, x_2, \cdots, x_k) = \frac{f(x_1, x_2, \cdots, x_k) - f(x_0, x_1, \cdots, x_{k-1})}{x_k - x_0}$$

特殊地，称函数本身为零阶差商，如 $f(x_0)$ 等。

与差分类似，差商也可以用列表的形式给出，如表 5－4 所示。

由表 5－4 可见，造差商表要比造差分表困难些，因为除了减法运算外，还要作除法运算。同时，也得了一个规律，某阶差商总是左边相邻两数之差与所用到的自变量两个端点之差的商。

表 5-4

x_i	$f(x_i)$	$f(x_i,x_{i+1})$	$f(x_i,x_{i+1},x_{i+2})$	$f(x_i,x_{i+1},x_{i+2},x_{i+3})$	$f(x_i,x_{i+1},x_{i+2},x_{i+3},x_{i+4})$
x_0	$f(x_0)$				
		$f(x_0,x_1)$			
x_1	$f(x_1)$		$f(x_0,x_1,x_2)$		
		$f(x_1,x_2)$		$f(x_0,x_1,x_2,x_3)$	
x_2	$f(x_2)$		$f(x_1,x_2,x_3)$		$f(x_0,x_1,x_2,x_3,x_4)$
		$f(x_2,x_3)$		$f(x_1,x_2,x_3,x_4)$	
x_3	$f(x_3)$		$f(x_2,x_3,x_4)$		
		$f(x_3,x_4)$			
x_3	$f(x_4)$				

2. 差商的重要性质

(1)常数的差商等于零。

(2)函数的和的差商等于各自差商的和。

设函数 $F(x)=f(x)+g(x)$,则

$$F(x_0,x_1)=\frac{F(x_1)-F(x_0)}{x_1-x_0}=\frac{f(x_1)+g(x_1)-f(x_0)+g(x_0)}{x_1-x_0}$$

$$=\frac{f(x_1)-f(x_0)}{x_1-x_0}+\frac{g(x_1)-g(x_0)}{x_1-x_0}$$

$$=f(x_0,x_1)+g(x_0,x_1)$$

(3)常数与函数乘积的差商等于常数与函数差商之积。

设 $F(x)=Af(x)$,则

$$F(x_0,x_1)=\frac{F(x_1)-F(x_0)}{x_1-x_0}=\frac{A[f(x_1)-f(x_0)]}{x_1-x_0}=Af(x_0,x_1)$$

(4)n 次多项式的 n 阶差商等于常数,高于 n 阶的差商等于零。

此性质只要证明多项式 $P(x)=x^k$ 的一阶差商是 $k-1$ 次多项式,再由性质(2)和(3)即可得证。事实上,

$$P(x_0,x)=\frac{x^k-x_0^k}{x-x_0}=x^{k-1}+x^{k-2}x_0+\cdots+x_0^{k-1}$$

(5)函数的任意阶差商可以通过各节点上的函数值线性表示。

对一阶差商,显然有

$$f(x_0,x_1)=\frac{f(x_1)-f(x_0)}{x_1-x_0}=\frac{f(x_1)}{x_1-x_0}+\frac{f(x_0)}{x_0-x_1}$$

而二阶差商为

$$f(x_0,x_1,x_2)=\frac{f(x_1,x_2)-f(x_0,x_1)}{x_2-x_0}$$
$$=\frac{1}{x_2-x_0}\left[\frac{f(x_1)}{x_1-x_2}+\frac{f(x_2)}{x_2-x_1}-\frac{f(x_0)}{x_0-x_1}-\frac{f(x_1)}{x_1-x_0}\right]$$
$$=\frac{f(x_0)}{(x_0-x_1)(x_0-x_2)}+\frac{f(x_1)}{(x_1-x_0)(x_1-x_2)}+\frac{f(x_2)}{(x_2-x_0)(x_2-x_1)}$$

依次可推得

$$f(x_0,x_1,x_2,\cdots,x_n)=\sum_{k=0}^{n}\frac{f(x_k)}{(x_k-x_0)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}$$

由于

$$\omega(x)=(x-x_0)(x-x_1)\cdots(x-x_n)$$

则

$$\omega'(x_k)=(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)$$

于是

$$f(x_0,x_1,\cdots,x_n)=\sum_{k=0}^{n}\frac{f(x_k)}{\omega'(x_k)}$$

从差商的这种表示形式可知差商与其自变量(节点)的排列次序无关,即任意改变节点顺序,差商值并不改变,这称为差商关于自变量的对称性。

从上式可以看出,每个节点对应于和式中的一项,它们的地位完全是平等的,节点的顺序改变无非是和式对应项的交换,并不改变其值。

此外,差商的上述和式使我们回想到拉格朗日插值公式。事实上,拉格朗日插值多项式可定为

$$P_n(x)=\sum_{k=0}^{n}\frac{\omega(x)f(x_k)}{\omega'(x_k)(x-x_k)}$$

(6)插值节点的函数值$f(x_k)$可以通过各阶差商表示出来。即$f(x_k)$可以由差商$f(x_0)$,$f(x_0,x_1)$,$f(x_0,x_1,x_2)\cdots,f(x_0,x_1,\cdots,x_n)$表示出来。

当$n=1$时

$$f(x_1)=f(x_0)+(x_1-x_0)f(x_0,x_1)$$

当$n=2$时

$$f(x_2)=f(x_1)+(x_2-x_1)f(x_1,x_2)$$

又

$$f(x_1,x_2)=f(x_0,x_1)+(x_2-x_0)\cdot f(x_0,x_1,x_2)$$

所以

$$f(x_2)=f(x_0)+(x_1-x_0)f(x_0,x_1)+(x_2-x_1)[f(x_0,x_1)+(x_2-x_0)f(x_0,x_1,x_2)]$$
$$=f(x_0)+(x_2-x_0)f(x_0,x_1)+(x_2-x_0)(x_2-x_1)f(x_0,x_1,x_2)$$

同理,可推得

$$f(x_n)=f(x_0)+(x_n-x_0)f(x_0,x_1)+(x_n-x_0)(x_n-x_1)f(x_0,x_1,x_2)+\cdots+$$
$$(x_n-x_0)(x_n-x_1)\cdots(x_n-x_{n-1})f(x_0,x_1,\cdots,x_n)$$

(三)差商、微商、差分三者之间的关系

1. 差商与微商的关系

差商与微商之间的关系可由下面的定理给出。

定理 若$f(x)$在(a,b)上n次可微,在$[a,b]$上的插值节点为$x_0,x_1,\cdots,x_n$,则存在ξ $(a<\xi<b)$,使得

$$f(x_0,x_1,\cdots,x_n)=\frac{1}{n!}f^{(n)}(\xi)$$

证明 考虑多项式

$$P(x)=f(x_0)+(x-x_0)f(x_0,x_1)+(x-x_0)(x-x_1)f(x_0,x_1,x_2)+\cdots+$$
$$(x-x_0)(x-x_1)\cdots(x-x_{n-1})f(x_0,x_1,\cdots,x_n)$$

若注意到上一段中得到的$f(x_k)$的表达式可知

$$P(x_k)=f(x_k)$$

现引入一辅助函数$\varphi(x)=f(x)-P(x)$,此函数显然n次可微,且在$[a,b]$内有$n+1$个不同的零点$x_0,x_1,x_2,\cdots,x_n$。由罗尔定理可知,在$[a,b]$内,$\varphi'(x)$至少有n个零点,$\varphi''(x)$至少有$n-1$个零点,如此下去,$\varphi^{(n)}(x)$在$[a,b]$内至少有一个零点,设此零点为$\xi,a<\xi<b$,则

$$\varphi^{(n)}(x)=f^{(n)}(x)-n!\cdot f(x_0,x_1,x_2,\cdots,x_n)$$

于是

$$f(x_0,x_1,x_2,\cdots,x_n)=\frac{1}{n!}f^{(n)}(\xi)$$

2. 差商与差分的关系

由于差分要求等距节点,所以差分与差商之间的关系亦应在等距节点之下考虑。

设$x_k=x_0+kh(k=0,1,2,\cdots,n)$,于是

一阶差商

$$f(x_0,x_1)=\frac{f(x_1)-f(x_0)}{x_1-x_0}=\frac{\Delta f(x_0)}{h}$$

$$f(x_1,x_2)=\frac{f(x_2)-f(x_1)}{x_2-x_1}=\frac{\Delta f(x_1)}{h}$$

二阶差商

$$f(x_0,x_1,x_2)=\frac{f(x_1,x_2)-f(x_0,x_1)}{x_2-x_0}$$

$$=\frac{1}{2h}\left[\frac{\Delta f(x_1)}{h}-\frac{\Delta f(x_0)}{h}\right]$$

$$=\frac{1}{2!\ h^2}\Delta^2 f(x_0)$$

依此推得 n 阶差商

$$f(x_0,x_1,x_2,\cdots,x_n)=\frac{1}{n!\ h^n}\Delta^n f(x_0)$$

3. 差分与微商的关系

同样，这也只是在等距节点的情况下考虑。其关系由如下定理来说明。

定理 若 $f(x)$ 在区间 (a,b) 上 n 次可微，那么在区间 (x_0,x_0+nh) 内有一点 ξ，使得 $\Delta^n f(x_0)=h^n f^{(n)}(\xi)$。

证明 由于

$$f(x_0,x_1,x_2,\cdots,x_n)=\frac{1}{n!\ h^n}\Delta^n f(x_0)=\frac{1}{n!}f^{(n)}(\xi)$$

所以

$$\Delta^n f(x_0)=h^n f^{(n)}(\xi)$$

由上式可以看出，若 h 很小，则差分是 h^n 阶小量，当步长减少 2 倍时，n 阶差分将减少 2^n 倍。

二、牛顿插值公式

由于具有 $n+1$ 个节点的次数不超过 n 次的插值多项式的唯一性，所以本节讨论的牛顿插值公式与前面的拉格朗日插值公式本质上是一样的，即它们是恒等的，只不过是表现形式的不同。正是由于这种表示形式的不同，使得牛顿插值公式比拉格朗日插值公式的计算更方便，这不仅表现在可以有差商与差分作为工具，而且当增加节点时，不像拉格朗日公式那样要全部重复计算。

牛顿插值公式分为不等距节点和等距节点公式两类。

（一）不等距节点的牛顿基本插值公式

这个公式实际上在本节讨论差商与微商的关系时已经给出了，即

$$P_n(x)=f(x_0)+(x-x_0)f(x_0,x_1)+(x-x_0)(x-x_1)f(x_0,x_1,x_2)+\cdots+$$
$$(x-x_0)(x-x_1)\cdots(x-x_{n-1})f(x_0,x_1,\cdots,x_n)$$

现讨论插值公式的余项。$P_n(x)$ 是一个 n 次插值多项式，对于插值区间 $[a,b]$ 上的任意一点 x，不防把它看作新的插值节点，即 $x_0,x_1,\cdots,x_n,x$。于是由差商的性质(6)可得

$$f(x)=f(x_0)+(x-x_0)f(x_0,x_1)+(x-x_0)(x-x_1)f(x_0,x_1,x_2)+\cdots+$$
$$(x-x_0)(x-x_1)\cdots(x-x_n)f(x_0,x_1,\cdots,x_n,x)$$

即

$$f(x)=P_n(x)+(x-x_0)(x-x_1)\cdots(x-x_n)f(x_0,x_1,\cdots,x_n,x)$$

于是,牛顿插值多项式的余项为

$$R_n(x)=f(x)-P_n(x)=(x-x_0)(x-x_1)\cdots(x-x_n)f(x_0,x_1,\cdots,x_n,x)$$

关于余项 $R_n(x)$ 有下面的余项定理。

定理 设 $f(x)$ 在 (a,b) 上 $n+1$ 次可微,$x_0,x_1,\cdots,x_n,x\in[a,b]$,则插值公式余项

$$R_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}\omega(x)$$

其中,$\omega(x)=\prod_{k=0}^{n}(x-x_k)$。

事实上,只要注意到差商与微商的关系,显然定理是成立的。如果 $f^{(n+1)}(\xi)$ 在 $[a,b]$ 上有上界 M,则可得余项式的估计式

$$|R_n(x)|\leqslant\frac{M}{(n+1)!}|\omega(x)|$$

例 5-4 已知函数 $f(x)$ 的某些对应值如下表,试用牛顿插值公式计算 $f(0.596)$ 的近似值,并估计误差。

x_k	0.4	0.55	0.65	0.80	0.90	1.05
$f(x_k)$	0.410 75	0.578 15	0.696 75	0.888 11	1.026 52	1.253 85

解 首先造差商表:

x_k	$f(x_k)$	一阶差商	二阶差商	三阶差商	四阶差商	五阶差商
0.40	0.410 75					
		1.116 0				
0.55	0.578 15		0.280 0			
		1.186 0		0.197 0		
0.65	0.696 75		0.358 8		0.033 4	
		1.275 7		0.213 7		−0.001 2
0.80	0.888 11		0.433 6		0.032 6	
		1.384 1		0.230 0		
0.90	1.026 52		0.525 6			
		1.515 5				
1.05	1.253 85					

由差商表看到,四阶差商接近相等,五阶差商接近于零,因此可用 4 次牛顿插值多项

式进行计算。这样，必须从表中选定5个插值节点。由于插值点 $x=0.596$，则选与 x 靠近的 x_0,x_1,x_2,x_3,x_4 为节点得

$$P_4(x)=0.410\ 75+1.116\ 0(x-0.40)+0.280\ 0(x-0.40)(x-0.55)+$$
$$0.197\ 0(x-0.40)(x-0.55)(x-0.65)+$$
$$0.033\ 4(x-0.4)(x-0.55)(x-0.65)(x-0.80)$$
$$P_4(0.596)=0.613\ 8\approx f(0.596)$$

计算误差为

$$|R_4(x)|=|f(x_0,x_1,x_2,x_3,x_4,x_5)\omega_5(0.596)|=0.001\ 2\omega_5(0.596)=6.7\times10^{-8}$$

（二）等距节点的牛顿前插及后插公式

当插值节点是等距的时候，利用差分与差商的关系，可以把牛顿基本插值多项式加以简化，就得到牛顿前插及后插公式。这是两个常用的插值多项式，它们分别用于计算开始节点及最后节点附近的函数值，所以也常称为表初及表末内插多项式。

1. 牛顿前插多项式

已知牛顿基本插值多项式

$$P_n(x)=f(x_0)+(x-x_0)f(x_0,x_1)+(x-x_0)(x-x_1)f(x_0,x_1,x_2)+\cdots+$$
$$(x-x_0)(x-x_1)\cdots(x-x_{n-1})f(x_0,x_1,\cdots,x_n)$$

当插值节点为等距时，若步长为 h，则有

$$x_k=x_0+kh\quad(k=0,1,2,\cdots,n)$$

现令 $x=x_0+th$，即 $t=\dfrac{1}{h}(x-x_0)$，又

$$f(x_0,x_1,\cdots,x_k)=\frac{1}{k!}\frac{1}{h^k}\Delta^k f(x_0)$$

将它们代入牛顿基本插值多项式即得牛顿前插多项式：

$$P_n(x)=P_n(x_0+th)=f(x_0)+\frac{t}{1!}\Delta f(x_0)+\frac{t(t-1)}{2!}\Delta^2 f(x_0)+\cdots+$$
$$\frac{t(t-1)(t-2)\cdots(t-n+1)}{n!}\Delta^n f(x_0)$$

依照差分与微商的关系，得到牛顿前插式的余项为

$$R_n(x)=R_n(x_0+th)=\frac{h^{n+1}}{(n+1)!}f^{(n+1)}(\xi)t(t-1)\cdots(t-n)$$

其中，$x_0<\xi<x_n$。

例5-5 对某液体实验得温度与黏度的关系如下表所示，试求该液体在23 ℃时的黏度值。

T/℃	20	22	24	26	28	30
μ	1.005 1	0.957 9	0.914 2	0.873 7	0.836 0	0.800 7

解 由所给数据作差分表：

T/℃	μ	$\Delta\mu$	$\Delta^2\mu$	$\Delta^3\mu$	$\Delta^4\mu$
20	1.005 1				
		−0.047 2			
22	0.957 9		0.003 5		
		−0.043 7		−0.000 3	
24	0.914 2		0.003 2		−0.000 1
		−0.040 5		−0.000 4	
26	0.873 7		0.002 8		0.000 0
		−0.037 7		−0.0004	
28	0.836 0		0.002 8		
		−0.035 3			
30	0.800 7				

由差分表可见 $\Delta^3\mu$ 已近于常数，所以取 3 次牛顿前插多项式进行计算：

$$P_3(T)=P_3(T_0+ht)=\mu_0+t\Delta\mu_0+\frac{1}{2!}t(t-1)\Delta^2\mu_0+\frac{1}{3!}t(t-1)(t-2)\Delta^3\mu_0$$

这里，插值点 $T=23$ ℃，$T_0=20$ ℃，$h=2$。

于是将 $t=\dfrac{23-20}{2}=1.5$ 代入上式得

$$P_3(23)=1.005\ 1-1.5\times0.047\ 2+\frac{1.5\times0.5}{2}\times0.003\ 5+\frac{1.5\times0.5\times0.5}{6}\times0.000\ 3$$

$$=0.935\ 6$$

计算误差为

$$|R_3(T)|=\left|\frac{t(t-1)(t-2)(t-3)}{4!}\Delta^4\mu_0\right|$$

$$=\frac{1}{24}|1.5\times0.5\times0.5\times1.5\times(-0.000\ 1)|$$

$$=2.34\times10^{-6}$$

2. 牛顿后插多项式

若把所给的函数表倒过来，换句话说从 x_n 点构造牛顿基本插值多项式，而不是从 x_0 点来构造 $P_n(x)$，则基本插值多项式有如下的形式：

$$P_n(x)=f(x_n)+(x-x_n)f(x_n,x_{n-1})+(x-x_n)(x-x_{n-1})f(x_n,x_{n-1},x_{n-2})+\cdots+$$
$$(x-x_n)(x-x_{n-1})\cdots(x-x_1)f(x_n,x_{n-1},x_0)$$

对于步长为 h 的等距节点有

$$x_{n-k}=x_n-kh\quad(k=0,1,2,\cdots n)$$

现作变量代换 $x = x_n + ht$，即 $t = \frac{1}{h}(x - x_n)$，由于

$$f(x_n, x_{n-1}, \cdots, x_{n-k}) = f(x_{n-k}, x_{n-k+1}, \cdots, x_{n-1}, x_n) = \frac{\Delta^k f(x_{n-k})}{k!\, h^k} = \frac{\nabla^k f(x_n)}{k!\, h^k}$$

所以

$$\begin{aligned} P_n(x) &= P_n(x_n + ht) \\ &= f(x_n) + \frac{t}{1!}\nabla f(x_n) + \frac{t(t+1)}{2!}\nabla^2 f(x_n) + \cdots + \frac{t(t+1)(t+2)\cdots(t+n-1)}{n!}\nabla^n f(x_n) \end{aligned}$$

且

$$R_n(x) = \frac{t(t+1)(t+2)\cdots(t+n)}{(n+1)!}h^{n+1} f^{(n+1)}(\xi)$$

其中，$x_0 < \xi < x_n$。

例 5-6 试求例 5-5 中的液体在 27 ℃时的黏度值。

解 由于 $T = 27$ ℃靠近表末，所以应采用牛顿后插多项式进行计算。注意此时的各阶差分是差分表中后斜线上的相应值。

这时 $T_n = 30$ ℃，$h = 2$，所以

$$t = \frac{27 - 30}{2} = -1.5$$

$$\begin{aligned} P_3(27) &= 0.800\,7 + (-1.5) \times (-0.035\,3) + \frac{1}{2} \times (-1.5) \times (-0.5) \times 0.002\,4 + \\ &\quad \frac{1}{6} \times (-1.5) \times (-0.5) \times 0.5 \times (-0.000\,4) \\ &= 0.854\,1 \end{aligned}$$

第四节 分段低次插值

用次数较高的插值多项式进行计算时，不仅由于高阶差商与差分会造成误差的扩散与积累，而且由于高次多项式振动频率高，在某些插值节点附近振幅也很大，导致所求近似值严重失真。而许多函数本身的图形就比较简单，用次数很高的多项式去逼近也没有必要。因此，在实际中，多采用低次多项式在区间 $[a, b]$ 上作分段插值，这样不仅计算简单，而且也会得到较理想的精度。

常用的分段低次插值有分段线性插值和分段抛物插值两种。

一、分段线性插值法

所谓分段线性插值，从图 5-2 上可见，就是用折线代替曲线，在每个小区间上，就是用直线来代替弧线。

对给定的插值节点 (x_i, y_i) $(i = 0, 1, 2, \cdots, n)$，由拉格朗日插值多项式的线性形式可

写出任意小区间$[x_i,x_{i+1}]$上的线性插值公式：

$$P_1(x)=\frac{x-x_{i+1}}{x_i-x_{i+1}}y_i+\frac{x-x_i}{x_{i+1}-x_i}y_{i+1}\quad(i=0,1,\cdots,n-1)$$

二、分段抛物插值法

抛物插值法就是用一组在区间$[x_{i-1},x_{i+1}](i=1,2,\cdots,n-1)$上首尾相连的抛物线（如图5-3所示）来代替函数$y=f(x)$的曲线。

对每个小区间$[x_{i-1},x_{i+1}]$可写出抛物插值公式：

$$P_2(x)=\frac{(x-x_i)(x-x_{i+1})}{(x_{i-1}-x_i)(x_{i-1}-x_{i+1})}y_{i-1}+\frac{(x-x_{i-1})(x-x_{i+1})}{(x_i-x_{i-1})(x_i-x_{i+1})}y_i+$$
$$\frac{(x-x_{i-1})(x-x_i)}{(x_{i+1}-x_{i-1})(x_{i+1}-x_i)}y_{i+1}\quad(i=1,2,\cdots,n-1)$$

分段插值的计算十分简单，首先判定被插值点x所属的小区间，然后根据选用的分段插值法进行计算。

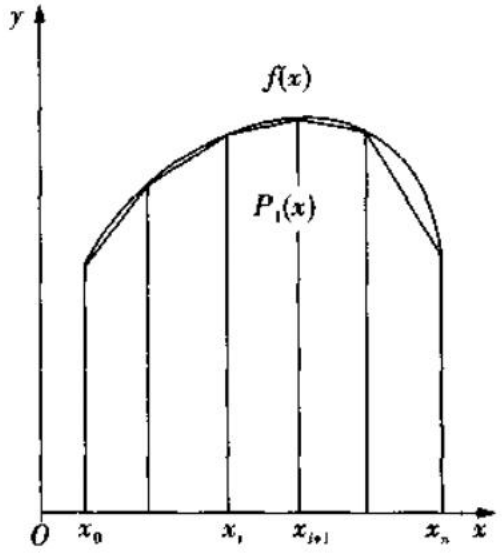


图5-2

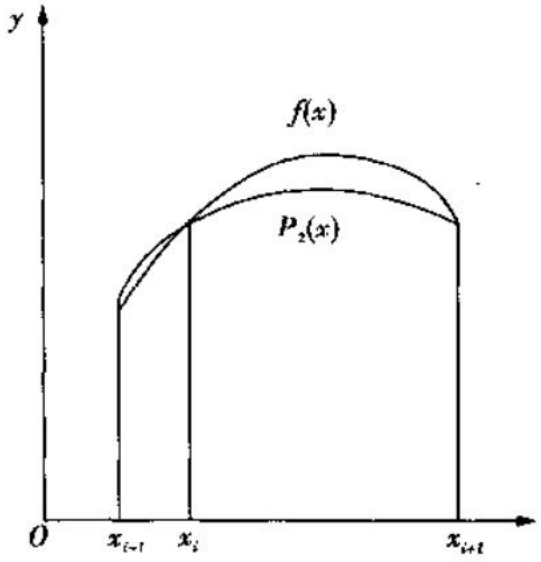


图5-3

第五节　三次样条函数插值

分段低次插值有许多优点。但是，这种插值方法只能保证各段曲线在连接点处的连续性，而不能保证它们在这些点上的光滑性，然而某些实际问题，既要求曲线连续，又要求曲线有连续的一阶乃至二阶导数，即要求变化率与曲率均连续。如船体、飞机等外型的曲线就是这样。本节所讨论的三次样条函数就能满足这种要求，它是在工程中应用较广泛的一类插值函数。

样条在物理上就是指在若干点上加载后能自然弯曲的条体。如曲线尺就属于这种富有弹性，能在一定的载荷下自然弯曲的条体。沿加载后的条体画下来的曲线就称为样条曲线，曲线所代表的函数称为样条函数。

一、三次样条函数的建立

定义 设在$[a,b]$上，函数$y=f(x)$在给定插值节点$a\leqslant x_0<x_1<\cdots<x_n\leqslant b$处的函数值$y_k=f(x_k)(k=0,1,2,\cdots,n)$为已知。若存在有函数$S(x)$满足以下条件：

(1)$S(x_k)=y_k(k=0,1,2,\cdots n)$；

(2)在节点x_k处具有连续的一阶和二阶导数，即

$$\begin{cases}S'(x_{k-0})=S'(x_{k+0})\\S''(x_{k-0})=S''(x_{k+0})\end{cases}\quad(k=1,2,\cdots,n-1)；$$

(3)在小区间$[x_k,x_{k+1}]$上，$S(x)$是不超过三次的多项式。

则称$S(x)$为$f(x)$在所给节点上的三次样条插值函数。

按照上述定义，可以推导出$S(x)$的计算公式。

设在节点x_k处$S(x)$的二阶导数为M_k，即

$$S''(x_k)=M_k\quad(k=0,1,2,\cdots,n)$$

由于假定了$S(x)$是每个小区间$[x_k,x_{k+1}]$上的三次多项式，所以$S''(x)$在$[x_k,x_{k+1}]$上一定是线性函数，于是有

$$S''(x)=\frac{x_{k+1}-x}{x_{k+1}-x_k}M_k+\frac{x-x_k}{x_{k+1}-x_k}M_{k+1}\quad(k=0,1,2,\cdots,n-1)$$

令$h_k=x_{k+1}-x_k$，则上式可写为

$$S''(x)=\frac{x_{k+1}-x}{h_k}M_k+\frac{x-x_k}{h_k}M_{k+1}\tag{5-1}$$

对式(5-1)积分得

$$S'(x)=-\frac{(x_{k+1}-x)^2}{2h_k}M_k+\frac{(x-x_k)^2}{2h_k}M_{k+1}+C_1\tag{5-2}$$

其中C_1是积分常数。

再对式(5-2)积分得

$$S(x)=\frac{(x_{k+1}-x)^3}{6h_k}M_k+\frac{(x-x_k)^2}{6h_k}M_{k+1}+C_1x+C_2\tag{5-3}$$

其中，C_2是积分常数。

代入已知条件$S(x_k)=y_k$，则

$$S(x_k)=y_k=\frac{h_k^2}{6}M_k+C_1x_k+C_2\tag{5-4}$$

$$S(x_{k+1})=y_{k+1}=\frac{1}{6}h_k^2M_{k+1}+C_1x_{k+1}+C_2\tag{5-5}$$

由式(5-5)减式(5-4)得

$$y_{k+1}-y_k=\frac{1}{6}h_k^2(M_{k+1}-M_k)+C_1h_k$$

则

$$C_1=\frac{1}{h_k}(y_{k+1}-y_k)-\frac{1}{6}h_k(M_{k+1}-M_k) \tag{5-6}$$

于是

$$C_2=y_k-\frac{h_k^2}{6}M_k-\frac{y_{k+1}-y_k}{h_k}x_k+\frac{h_k}{6}(M_{k+1}-M_k)x_k$$

$$C_1x+C_2=y_k-\frac{h_k^2}{6}M_k+\frac{y_{k+1}-y_k}{h_k}(x-x_k)-\frac{h_k}{6}(M_{k+1}-M_k)(x-x_k) \tag{5-7}$$

将式(5-7)代入式(5-3)整理得

$$S(x)=\frac{(x_{k+1}-x)^3}{6h_k}M_k+\frac{(x-x_k)^3}{6h_k}M_{k+1}+\left(y_k-\frac{h_k^2}{6}M_k\right)\frac{(x_{k+1}-x)}{h_k}+$$

$$\left(y_{k+1}-\frac{h_k^2}{6}M_{k+1}\right)\frac{x-x_k}{h_k}\quad(k=0,1,\cdots,n-1) \tag{5-8}$$

式(5-8)就是三次样条插值公式。只要能定出 $M_k(k=0,1,\cdots,n-1)$ 的值，就可以利用该式进行计算。下面给出确定 M_k 的计算方法。

二、三次样条函数的计算

1. 确定计算 M_k 的数学模型

根据假设，函数 $S(x)$ 应在各插值节点有连续的一阶导数，所以一阶导数的左右极限应该满足

$$S'(x_{k-0})=S'(x_{k+0})$$

如图 5-4 所示，x_{k-0} 与 x_{k+0} 分别处于小区间 $[x_{k-1},x_k]$ 与 $[x_k,x_{k+1}]$ 内。而 $S'(x)$ 在这两个小区间上的表达式分别为：

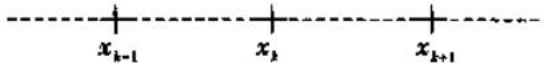


图 5-4

在 $[x_{k-1},x_k]$ 上

$$S'(x)=-\frac{(x_k-x)^2}{2h_{k-1}}M_{k-1}+\frac{(x-x_{k-1})^2}{2h_{k-1}}M_k+\frac{y_k-y_{k-1}}{h_{k-1}}-\frac{h_{k-1}}{6}(M_k-M_{k-1})$$

在 $[x_k,x_{k+1}]$ 上

$$S'(x)=-\frac{(x_{k+1}-x)^2}{2h_k}M_k+\frac{(x-x_k)^2}{2h_k}+\frac{y_{k+1}-y_k}{h_k}-\frac{h_k}{6}(M_{k+1}-M_k)$$

于是可得出

$$S'(x_{k-0})=\frac{h_{k-1}}{2}M_k-\frac{h_{k-1}}{6}(M_k-M_{k-1})+\frac{y_k-y_{k-1}}{h_{k-1}}$$

$$S'(x_{k+0})=-\frac{h_k}{2}M_k-\frac{1}{6}h_k(M_{k+1}-M_k)+\frac{y_{k+1}-y_k}{h_k}$$

则有

$$6S'(x_{k-0})=2h_{k-1}M_k+h_{k-1}M_{k-1}+6f(x_{k-1},x_k)$$

$$6S'(x_{k+0})=-2h_kM_k-h_kM_{k+1}+6f(x_k,x_{k+1})$$

这样

$$-2h_kM_k-h_kM_{k+1}+6f(x_k,x_{k+1})=2h_{k-1}M_k+h_{k-1}M_{k-1}+6f(x_{k-1},x_k)$$

整理得

$$h_{k-1}M_{k-1}+2(h_k+h_{k-1})M_k+h_kM_{k+1}=6[f(x_k,x_{k+1})-f(x_{k-1},x_k)]$$

又

$$h_k+h_{k-1}=x_{k+1}-x_k+x_k-x_{k-1}=x_{k+1}-x_{k-1}$$

$$\frac{h_{k-1}}{h_k+h_{k-1}}M_{k-1}+2M_k+\frac{h_k}{h_k+h_{k-1}}M_{k+1}=6\frac{f(x_k,x_{k+1})-f(x_{k-1},x_k)}{x_{k+1}-x_{k-1}}$$

即

$$\frac{h_{k-1}}{h_k+h_{k-1}}M_{k-1}+2M_k+\frac{h_k}{h_k+h_{k-1}}M_{k-1}=6f(x_{k-1},x_k,x_{k+1})$$

令

$$u_k=\frac{h_{k-1}}{h_k+h_{k-1}},\quad v_k=\frac{h_k}{h_k+h_{k-1}},\quad d_k=6f(x_{k-1},x_k,x_{k+1})$$

故

$$u_kM_{k-1}+2M_k+v_kM_{k+1}=d_k\quad(k=1,2,\cdots,n-1)$$

二阶导数 M_k 在力学上解释为梁在 x_k 处的弯矩，所以上式也称为三弯矩方程。由于上式是一个有 $n+1$ 个未知数的 $n-1$ 个方程组成的方程组，所以还需要补充两个条件才能定解。这两个条件通常由插值区间的端点处得到，也叫做边值条件，常见的边值条件有以下两种情形。

(1)已知插值区间两端点的二阶导数

$$M_0=S''(x_0),\quad M_n=S''(x_n)$$

在方程组中，M_0 与 M_n 仅分别出现在第一个与最后一个方程中，按边值条件，现在它们的值为已知。于是把 M_0 和 M_n 分别代入第一个与最后一个方程，消去两个未知数，则得方程组为

$$\begin{bmatrix} 2 & v_1 & & & \\ u_2 & 2 & v_2 & & \\ & \ddots & \ddots & \ddots & \\ & & u_{n-2} & 2 & v_{n-2} \\ & & & u_{n-1} & 2 \end{bmatrix} \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{bmatrix} = \begin{bmatrix} d_1 - u_1 M_0 \\ d_2 \\ \vdots \\ d_{n-2} \\ d_{n-1} - v_{n-1} M_n \end{bmatrix}$$

这是一个具有 $n-1$ 个未知数 $M_k(k=1,2,\cdots,n-1)$ 的 $n-1$ 个方程组成的方程组，且其系数矩阵为严格对角优势矩阵，所以其解是唯一存在的，可用追赶法求解。

(2)已知插值区间两端点的一阶导数

$$m_0 = S'(x_0), \quad m_n = S'(x_n)$$

应用这个条件，并注意 $x_0 \in [x_0, x_1]$，$x_n \in [x_{n-1}, x_n]$，则可得到两个方程

$$2M_0 + M_1 = \frac{6}{h_0}[f(x_0, x_1) - m_0]$$

$$M_{n-1} + 2M_n = \frac{6}{h_{n-1}}[m_n - f(x_{n-1}, x_n)]$$

把这两个方程补充到前面的方程组中，就得到方程组

$$\begin{bmatrix} 2 & 1 & & & \\ u_1 & 2 & v_1 & & \\ & u_2 & 2 & v_2 & \\ & & \ddots & \ddots & \ddots \\ & & & u_{n-1} & 2 & v_{n-1} \\ & & & & 1 & 2 \end{bmatrix} \begin{bmatrix} M_0 \\ M_1 \\ M_2 \\ \vdots \\ M_{n-1} \\ M_n \end{bmatrix} = \begin{bmatrix} \frac{6}{h_0}[f(x_0, x_1) - m_0] \\ d_1 \\ d_2 \\ \vdots \\ d_{n-1} \\ \frac{6}{h_{n-1}}[m_n - f(x_{n-1}, x_n)] \end{bmatrix}$$

此方程组的系数矩阵仍是一个严格对角优势的三对角矩阵，同样，可用追赶法求解。

当 $S'(x_0) = S'(x_n) = 0$ 时，称为固定边界。

2. 样条函数的插值计算

当 $M_k(k=0,1,\cdots,n)$ 确定之后，即可进行样条函数的插值计算。首先判定被插值点 x 所属的小区间 $[x_k, x_{k+1}]$，然后利用 $S(x)$ 进行插值计算。显然，样条函数插值仍为分段插值，在不同的小区间上计算公式的参量不同。

综上所述，样条插值分两步完成：

第一步，确定样条函数。由给定插值节点 $(x_k, y_k)(k=0,1,\cdots,n)$，通过 $u_i = \frac{h_{i-1}}{h_i + h_{i-1}}$，$v_i = \frac{h_i}{h_i + h_{i-1}}$，$d_i = 6f(x_{i-1}, x_i, x_{i+1})$，$i = 1,2,\cdots,n-1$，求出 u_i, v_i, d_i。再根据边值条件求解相应的方程组定出 $M_0, M_1, M_2, \cdots, M_n$。

第二步，进行插值计算。判定出被插值点所属的小区间，再利用 $S(x)$ 在该小区间上的表达式进行计算，即求得被插值点 x 处的函数值 $f(x)$。

例 5－7　某液相反应 $A \longrightarrow B + C$，实验测得反应物 A 的浓度 c_A 随时间 t 的变化情况为：

t /min	0	0.2	0.6	1.0	3.0	5.0	10.0
$c_A/(g \cdot L^{-1})$	5.19	3.77	2.30	1.57	0.8	0.25	0.094

并由图解法求得 $y_0' = \left.\dfrac{dc_A}{dt}\right|_{t=0} = -9.45$，$y_6' = \left.\dfrac{dc_A}{dt}\right|_{t=10.0} = 0$，试用三次样条插值函数计算 $t = 0.1, 0.4, 2.0, 4.0$ min 下的 c_A 值。

解　首先造差商表并求 d_k, u_k, v_k。

X_k	Y_k	$f(x_{k-1}, x_k)$	$f(x_{k-1}, x_k, x_{k+1})$	d_k
0	5.19			
		−7.1		
0.2	3.77		5.708	34.25
		−3.675		
0.6	2.30		2.312 5	13.875
		−1.825		
1.0	1.57		0.6	3.6
		−0.385		
3.0	0.8		0.027 5	0.165
		−0.275		
5.0	0.25		0.034 83	0.209
		−0.031 2		
10.0	0.094			

由

$$u_k = \frac{h_{k-1}}{h_k + h_{k-1}}, \quad v_k = \frac{h_k}{h_k + h_{k-1}} \quad (k = 0, 1, 2, 3, 4, 5)$$

求得

$$u_1 = 0.333, \quad u_2 = 0.5, \quad u_3 = 0.167, \quad u_4 = 0.5, \quad u_5 = 0.286;$$

$$v_1 = 0.667, \quad v_2 = 0.5, \quad v_3 = 0.833, \quad v_4 = 0.5, \quad v_5 = 0.714$$

代入边值条件 $y_0' = -9.45$，$y_6' = 0$ 得

$$\begin{bmatrix} 2 & 1 & & & & & \\ 0.333 & 2 & 0.667 & & & & \\ & 0.5 & 2 & 0.5 & & & \\ & & 0.167 & 2 & 0.833 & & \\ & & & 0.5 & 2 & 0.5 & \\ & & & & 0.286 & 2 & 0.714 \\ & & & & & 1 & 2 \end{bmatrix} \begin{bmatrix} M_0 \\ M_1 \\ M_2 \\ M_3 \\ M_4 \\ M_5 \\ M_6 \end{bmatrix} = \begin{bmatrix} 70.5 \\ 34.25 \\ 13.875 \\ 3.60 \\ 0.165 \\ 0.209 \\ 0.037\ 44 \end{bmatrix}$$

由追赶法求得

$$M_0 = 29.803, \quad M_1 = 10.893\ 9, \quad M_2 = 3.804\ 7, \quad M_3 = 1.637\ 5,$$
$$M_4 = -0.372\ 7, \quad M_5 = 0.183\ 9, \quad M_6 \approx -0.073\ 25$$

再进行插值计算可得

$$S(0.1) = 4.378\ 3\ (\text{g/L}), \quad S(0.4) \approx 2.888\ 2\ (\text{g/L}),$$
$$S(2.0) = 0.868\ 8\ (\text{g/L}), \quad S(4.0) = 0.073\ 25\ (\text{g/L})$$

第六节　埃尔米特插值多项式

在不少问题中，求取插值多项式时，不仅要求在节点 $x_0, x_1, \cdots, x_n$ 处函数值要满足

$$P(x_k) = f(x_k) \quad (k = 0, 1, 2, , n)$$

而且还要求在各节点处若干阶导数相等，即

$$P^{(r)}(x_k) = f^{(r)}(x_k) \quad (r = 0, 1, 2, \cdots; k = 0, 1, \cdots, n)$$

其中最常见的是要求在各节点处函数及其一阶导数相等，即

$$P(x_k) = f(x_k), \quad P'(x_k) = f'(x_k) \quad (k = 0, 1, 2, \cdots, n)$$

上式的几何意义是，要求曲线 $y = P(x)$ 和曲线 $y = f(x)$ 不但都经过 n 个共同点 (x_k, y_k)，而且在这些点处的切线相同。当然，一般来说，这样的插值多项式比起前面所讨论的插值多项式要好一些。

事实上，上述问题可叙述为：设给定函数 $y = f(x)$ 在 $n+1$ 个不同的点 $x_0, x_1 \cdots, x_n$ 处的函数 $y_0, y_1 \cdots, y_n$ 和导数 $y_0', y_1' \cdots, y_n'$，求一个多项式 $P(x)$，使 $P(x_i) = y_i, P'(x_i) = y_i'$ $(i = 0, 1, 2, \cdots, n)$。

可以看出，这里给出了 $2n+2$ 个条件，因而多项式 $P(x)$ 一般不超过 $2n+1$ 次。由于 $P(x_i) = y_i$，则可作出拉格朗日多项式 $F_n(x)$，它是 n 次多项式，满足 $F(x_i) = y_i$，通常它不满足 $F_n'(x_i) = y_i'$，但可以设想要找的 $P(x)$ 总可以表示为

$$P(x) = F_n(x) + Q(x)$$

上式中的 $Q(x)$，其次数应不超过 $2n+1$，且在节点处取零值，亦即 $Q(x_i) = 0$，于是 $Q(x)$ 又可以写为

$$Q(x)=\Phi_n(x)\cdot(x-x_0)(x-x_1)\cdots(x-x_n)=\Phi_n(x)\cdot\omega(x)$$

为了确定 $\Phi_n(x)$，就要利用条件 $P'(x_i)=y_i'$，于是

$$P'(x)=F_n'(x)+\Phi_n'(x)\omega(x)+\Phi_n(x)\omega'(x)$$

由于 $\omega(x_i)=0$，所以

$$P'(x_i)=F_n'(x_i)+\Phi_n(x_i)\omega'(x_i)=y_i'$$

于是

$$\Phi_n(x_i)=\frac{y_i'-F_n'(x_i)}{\omega'(x_i)}\quad(i=0,1,2,\cdots,n)$$

该式表明，特定的多项式 $\Phi_n(x)$ 在 $n+1$ 个节点处的取值已知。利用拉格朗日插值多项式可得 $\Phi_n(x)$ 为

$$\Phi_n(x)=\sum_{k=0}^{n}\frac{\omega(x)}{\omega'(x_k)(x-x_k)}\Phi_n(x_k)$$

即

$$\Phi_n(x)=\sum_{k=0}^{n}\frac{\omega(x)}{\omega'(x_k)(x-x_k)}\frac{y_k'-F_n'(x_k)}{\omega'(x_k)}$$

由于 $\Phi_n(x)$ 是次数不超过 n 的多项式，所以上式是精确成立的。若利用拉格朗日插值多项式的基函数表示，则上式可写为

$$\Phi_n(x)=\sum_{k=0}^{n}L_k(x)\cdot\frac{y_k'-F_n'(x_k)}{\omega'(x_k)}$$

故

$$P(x)=\sum_{k=0}^{n}y_kL_k(x)+\omega(x)\sum_{k=0}^{n}L_k(x)\frac{y_k'-F_n'(x_k)}{\omega'(x_k)}$$

此式即是按函数及其一阶导数进行插值的插值多项式，通常称为埃尔米特插值多项式。

为编制电子计算机程序的方便，上式可整理为

$$P(x)=\sum_{k=0}^{n}b_k^2[(x-x_k)(y_k'-2a_ky_k)+y_k]$$

式中

$$b_k=\prod_{\substack{j=0\\j\neq k}}^{n}\frac{x-x_j}{x_k-x_j}=L_k(x),\quad a_k=\sum_{\substack{j=0\\j\neq k}}^{n}\frac{1}{x_k-x_j}$$

由于当节点数 n 太大时，多项式的方次将很高，这不仅计算麻烦，而且受计算机容量的限制，所以 $n+1$ 个节点的埃尔米特插值多项式通常不单独使用，往往采用分段插值方法。

现考察一下只有两个节点的情形，即要在区间 $[x_0,x_1]$ 上作一个插值多项式 $P(x)$，使其满足

$$P(x_0)=y_0,\quad P(x_1)=y_1;$$
$$P'(x_0)=y_0',\quad P'(x_1)=y_1'$$

显然，此时 $n=1$，$P(x)$ 是三次多项式，于是

$$P(x)=\sum_{k=0}^{1}b_k^2[(x-x_k)(y_k'-2a_ky_k)+y_k]$$

式中，$b_0=\dfrac{x-x_1}{x_0-x_1}$，$b_1=\dfrac{x-x_0}{x_1-x_0}$，$a_0=\dfrac{1}{x_0-x_1}$，$a_1=\dfrac{1}{x_1-x_0}$。

于是可得

$$P(x)=\left(\frac{x-x_1}{x_0-x_1}\right)^2\left[(x-x_0)\left(y_0'-\frac{2y_0}{x_0-x_1}\right)+y_0\right]+$$
$$\left(\frac{x-x_0}{x_1-x_0}\right)^2\left[(x-x_1)\left(y_1'-\frac{2}{x_1-x_0}y_1\right)+y_1\right]$$

整理得

$$P(x)=\left(\frac{x-x_0}{x_1-x_0}\right)^2\left[1-\frac{2(x-x_1)}{x_1-x_0}\right]y_1+\left(\frac{x-x_1}{x_0-x_1}\right)^2\left[1-\frac{2(x-x_0)}{x_0-x_1}\right]y_0+$$
$$(x-x_1)\left(\frac{x-x_0}{x_1-x_0}\right)^2y_1'+(x-x_0)\left(\frac{x-x_1}{x_0-x_1}\right)^2y_0'$$

这样，对于任意的小区间$[x_k,x_{k+1}]$，埃尔米特插值多项式可写为

$$P(x)=\left(\frac{x-x_k}{x_{k+1}-x_k}\right)^2\left[1-\frac{2(x-x_{k+1})}{x_{k+1}-x_k}\right]y_{k+1}+\left(\frac{x-x_{k+1}}{x_k-x_{k+1}}\right)^2\left[1-\frac{2(x-x_k)}{x_k-x_{k+1}}\right]y_k+$$
$$(x-x_{k+1})\left(\frac{x-x_k}{x_{k+1}-x_k}\right)^2y_{k+1}'+(x-x_k)\left(\frac{x-x_{k+1}}{x_k-x_{k+1}}\right)^2y_k'$$

利用此式即可编制任意两相邻节点 x_k 和 x_{k+1} 的埃尔米特插值多项式的计算机程序。

例 5-8 已知某函数的两插值点 $x_0=0.1$ 和 $x_1=0.3$ 的函数值及导数值分别为 $y_0=0.099\ 83$，$y_1=0.295\ 5$，$y_0'=0.955\ 0$，$y_1'=0.955\ 3$，试求其在 $x=0.25$ 处的函数值。

解 由于 $x=0.25$，则

$$P(0.25)=\left(\frac{0.25-0.1}{0.3-0.1}\right)^2\left[1-\frac{2(0.25-0.3)}{0.3-0.1}\right]\times0.295\ 5+$$
$$\left(\frac{0.25-0.3}{0.1-0.3}\right)^2\left[1-\frac{2(0.25-0.1)}{0.1-0.3}\right]\times0.099\ 83+$$
$$(0.25-0.3)\left(\frac{0.25-0.1}{0.3-0.1}\right)^2\times0.955\ 3+$$
$$(0.25-0.1)\left(\frac{0.25-0.3}{0.1-0.3}\right)^2\times0.955$$
$$=0.247\ 4$$

习 题

1. 已知

x_i	1.690 0	1.988 1	2.016 4
$\sqrt{x_i}$	1.400 0	1.410 0	1.420 0

试分别用线性插值和抛物差值求$\sqrt{2}$，并估计误差，比较其准确程度。

2. 已知丙苯黏度在40～75 ℃范围内随着温度的变化数据如下：

T/℃	40	45	55	75
μ/(mPa·S)	0.68	0.64	0.56	0.45

试用拉格朗日二次插值公式和三次插值公式求丙苯在50 ℃和70 ℃时的黏度。

3. 利用牛顿插值多项式由下表中数据确定在20、30、40、50 ℃下氨蒸气的蒸气压。

T/℃	21.10	26.65	32.20	37.75	43.30	48.85	54.40
p/MPa	12.63	14.99	17.70	20.77	24.21	28.07	32.37

4. 已知丙烯的饱和蒸气压数据如下：

T/℃	-28.9	-12.2	4.4	21.1	37.8
p/atm	2.2	3.9	6.6	10.3	15.4

试用牛顿插值多项式求在-22 ℃和20 ℃时丙稀的饱和蒸汽压。

5. 实验测得乙炔的摩尔热容c_p与温度T的关系为：

T/K	300	400	500	600	700	800	900
c_p/(J·mol^{-1}·K^{-1})	41.382	46.273	50.703	54.507	57.768	60.652	63.118

现已知$\left.\frac{dc_p}{dT}\right|_{T=300\text{ K}}=0.050\ 04$，$\left.\frac{dc_p}{dT}\right|_{T=900\text{ K}}=0.022\ 35$。试用三次样条插值函数求温度为350、480、660 K时乙炔的c_p值。

6. 已知节点数据如下：

x	0	1	2	3
$f(x)$	0	0	0	0

试求当边界条件分别为：(1)$M_0 = M_3 = 1$；(2)$f'(0) = 0, f'(3) = 1$ 时三次样条插值函数的表达式，并求 x 为 0.5、1.5 和 2.5 时的函数值。

7. 已知函数 $y = f(x)$ 在两插值节点 $x_0 = 0.3, x_1 = 0.5$ 处的函数值和一阶导数分别为 $y_0 = 0.295\ 52, y_1 = 0.479\ 43, y_0' = 0.955\ 34, y_1' = 0.877\ 58$。试利用埃尔米特插值公式求 $x = 0.35$处的函数值。

第六章　函数的多项式逼近

函数的逼近就是寻找一条连续曲线$f(x)$逼近空间中的几何点$(x_i,y_i)(i=0,1,2,\cdots,n)$，故也称为曲线拟合。

化学工程中所采用的大量经验公式和半经验公式都是通过函数的逼近求得的。例如传递工程中的准数关联式，化工热力学中的热容量与温度间的关系式、蒸气压方程、状态方程，化工动力学中的速率表达式等。

由函数的插值方法的讨论可知，插值法是一种用函数$P_n(x)$来逼近空间几何点$(x_i,y_i)(i=0,1,2,\cdots,n)$的实用有效方法。但是插值法要求的条件比较严格，即要求在$[a,b]$内的$n+1$个节点$(x_i,y_i)(i=0,1,2,\cdots,n)$上插值函数$P_n(x)$与被插值函数$f(x)$精确地相等，也就是在$f(x_i)=P_n(x_i)(i=0,1,2,\cdots,n)$的条件下构造插值函数

$$P_n(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n$$

这就给改进插值函数的精度造成了一定的困难。

为了提高近似函数的精度，人们提出了在多种意义下函数逼近的概念，并提出了最佳逼近问题。所谓最佳逼近，通常指的是最佳一致逼近与最佳平方逼近。至于一致逼近的方法，读者可参阅有关参考书，这里主要讨论在化学工程中应用较广泛的最佳平方逼近。为了这种需要，给出了一些必要的数学概念，主要是内积与正交性的概念。

第一节　内积

一、线性空间

1. 线性空间的定义

为了引入线性空间的定义，先给出映射的概念。

(1)映射：给定两个集合M和N，如果又给定一个法则f，使M中每一个元素a在f作用下都对应于N中一个唯一确定的元素，记作$f(a)$，则称f是集合M到集合N的一个映

射，且$f(a)$称为a的对象，a称为$f(a)$的一个原象。

(2)一一映射：若集合M到集合N的一个映射f满足

(i)是单一的，即M中两个不同元素在f作用下，对应于N中的两个不同元素；

(ii)是满的，即N中任一个元素都是M中某一个元素在f作用下的象。

则称f是M到N的一一映射。

(3)线性空间：设V是一个非空集合，K是一个数域，又设

(i)在V中定义了一种运算，称为加法，即对V中任意两个元素α与β，都按某一法则对应于V内唯一确定的一个元素，记为$\alpha+\beta$；

(ii)在V中定义了一种运算，称为数乘，即对V中任意元素α和数域K中任意数k，都按某一法则对应于V内唯一确定的一个元素，记为$k\alpha$。

则称V是数域K上的一个线性空间。

2. 线性空间的性质

(1)交换律：$\alpha+\beta=\beta+\alpha$。

(2)结合律：$\alpha+(\beta+\gamma)=(\alpha+\beta)+\gamma$。

(3)存在一个元素$O\in V$，使对一切$\alpha\in V$，有$\alpha+O=\alpha$，此元素O称为V的零元素。

(4)对任意$\alpha\in V$，都存在$\beta\in V$，使得$\alpha+\beta=0$，并称β为α的一个负元素，即$\alpha=-\beta$。

(5)对任意$k\in K,\alpha,\beta\in V$，有$k(\alpha+\beta)=k\alpha+k\beta$。

(6)对任意$k,l\in K,\alpha\in V$有$(kl)\alpha=k(l\alpha)=l(k\alpha)$。

(7)对任意$l,k\in V,\alpha\in V$有$(k+l)\alpha=k\alpha+l\alpha$。

例6-1 证明n维向量空间是线性空间。

证明 事实上，设X,Y为任意两个向量

$$X=(x_1,x_2,\cdots,x_n),\quad Y=(y_1,y_2,\cdots,y_n)$$

则

$$X+Y=(x_1+y_1,x_2+y_2,\cdots,x_n+y_n),\quad KX=(kx_1,kx_2,\cdots,kx_n)$$

即n维向量空间是线性空间。

例6-2 证明n次多项式的全体不构成线性空间。

证明 因为$f(x)=x^n+x,g(x)=x-x^n$都属于n次多项式，但$f(x)+g(x)=2x$不属于n次多项式，所以n次多项式的全体不构成线性空间。

当线性空间定义中的数域K为实数域R时，线性空间V就是实数域中的线性空间。以下引入的线性空间中内积的概念都是在实数域中的线性空间。

二、线性空间的内积

1. 内积的定义

设E为线性空间，对于实数域R来说，若元素$X,Y\in E$，与实数(x,y)相对应，并满足以下条件：

(1)非负性,即对于 $X \in E$,恒有 $(X,X) \geqslant 0$,且仅当 $X=0$ 时等式成立;

(2)可交换性,即有 $(X,Y)=(Y,X)$;

(3)齐次性,即对于 $X,Y \in E$,λ 为实数,恒有

$$(\lambda X,Y)=\lambda(X,Y);$$

(4)分配性,即对于 X、Y、$Z \in E$,恒有

$$(X,Y+Z)=(X,Y)+(X,Z)。$$

则称在空间 E 上定义了内积 (X,Y)。

2. 几种常见的内积形式

(1)n 维向量空间的内积。若 $X,Y \in \mathrm{R}_n$,$X=(x_1,x_2,\cdots,x_n)$,$Y=(y_1,y_2,\cdots,y_n)$,则

$$(X,Y)=\sum_{i=1}^{n} x_i y_i$$

即向量的内积等于向量对应坐标的积之和。

(2)连续函数空间中的内积。设 $F_{[a,b]}$ 表示在 $[a,b]$ 上连续函数组成的空间,即

$$F_{[a,b]}=\{\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x),\cdots\},$$

则任意两个函数的内积用它们的乘积在 $[a,b]$ 上的积分来表示,即

$$(\varphi_j(x),\varphi_k(x))=\int_a^b \varphi_j(x)\varphi_k(x)\mathrm{d}x。$$

(3)离散函数(列表函数)的内积。若函数 $\varphi_k(x)$ $(k=1,2,\cdots,n)$ 只在 $[a,b]$ 中的某些点 x_i $(i=1,2,\cdots,m)$ 处取得数值,那么,任意两个函数 $\varphi_j(x)$ 和 $\varphi_k(x)$ 的内积为:

$$(\varphi_j(x),\varphi_k(x))=\sum_{i=1}^{m} \varphi_j(x_i)\varphi_k(x_i) \quad (k,j=1,2,\cdots,n)。$$

(4)带权情形下的内积。有时,为进一步反映出随着 x 的不同,函数 $\varphi_j(x)$ 与 $\varphi_k(x)$ 在地位上的差异(即重要性上的差异),可适当地引入一个函数 $\rho(x)>0$ 来描述这种差异,这样的函数 $\rho(x)$ 称为权函数,这种引入权函数的内积即是所谓带权情形下的内积。

对于连续函数

$$(\varphi_j(x),\varphi_k(x))=\int_a^b \rho(x)\varphi_j(x)\varphi_k(x)\mathrm{d}x$$

对于离散函数

$$(\varphi_j(x),\varphi_k(x))=\sum_{i=1}^{m} \rho(x_i)\varphi_j(x_i)\varphi_k(x_i)$$

特别地,当 $\rho(x)=1$ 时,即为不带权的内积。

第二节　正交多项式

一、函数的正交性和线性无关

1. 函数的正交性

从空间解析几何的知识可知，一条直线的方向是由它的3个方向数 x,y,z 所决定的，两个方向正交的条件是

$$x_1x_2+y_1y_2+z_1z_2=0$$

在 n 维空间中，每条直线是由它的 n 个方向数所决定的，两个方向正交的条件是

$$\sum_{i=1}^{n}x_iy_i=0$$

对区间 $[a,b]$ 上的两个函数 $f(x),g(x)$，若满足

$$(f,g)=\int_a^b\rho(x)f(x)g(x)\mathrm{d}x=0$$

就称函数 $f(x)$ 和 $g(x)$ 在区间 $[a,b]$ 上带权 $\rho(x)$ 正交。其中，$\rho(x)$ 为权函数。

若有一族函数 $\varphi_0(x),\varphi_1(x),\cdots,\varphi_i(x)\cdots$，假定每一个 $\varphi_i(x)$ 在 $[a,b]$ 上都连续，且不恒等于零，当满足

$$(\varphi_j(x),\varphi_k(x))=\int_a^b\rho(x)f(x)g(x)\mathrm{d}x=\begin{cases}0, & j\neq k\\ A_k>0, & j=k\end{cases}$$

时，称函数族 $\{\varphi_i(x)\}$ 为 $[a,b]$ 上的带权 $\rho(x)$ 的正交函数族。特别地，当 $A_k=1$ 时，称为标准正交函数族。

若每一个 $\varphi_i(x)$ 均为多项式，则 $\{\varphi_i(x)\}$ 称为 $[a,b]$ 上的正交多项式序列，并称 $\varphi_n(x)$ 为区间 $[a,b]$ 上的 n 次正交多项式。

2. 函数的线性无关

定义 若 $[a,b]$ 上的连续函数 $\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)$ 的线性组合关系式

$$a_0\varphi_0(x)+a_1\varphi_1(x)+\cdots+a_n\varphi_n(x)=0$$

仅当 $a_0=a_1=\cdots=a_n=0$ 时才成立，则称函数 $\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)$ 线性无关，否则为线性相关。

若函数族 $\{\varphi_i(x)\}(i=0,1,2,\cdots)$ 中，任何有限个函数组成的集合线性无关，则称函数族 $\{\varphi_i(x)\}$ 为线性无关的函数族。

由函数的正交性和线性无关的定义可得如下定理。

定理 在 $[a,b]$ 上带权 $\rho(x)$ 正交的函数族是线性无关函数族。

为了证明这个定理，先证明下面的引理。

引理 函数 $\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)$ 在 $[a,b]$ 上线性无关的充分必要条件是它们的克莱姆行列式 $G\neq0$，即

$$G_n=\begin{bmatrix}(\varphi_0,\varphi_0) & (\varphi_0,\varphi_1) & \cdots & (\varphi_0,\varphi_n)\\ (\varphi_1,\varphi_0) & (\varphi_1,\varphi_1) & \cdots & (\varphi_1,\varphi_n)\\ \vdots & \vdots & & \vdots\\ (\varphi_n,\varphi_0) & (\varphi_n,\varphi_1) & \cdots & (\varphi_n,\varphi_n)\end{bmatrix}\neq0$$

证明 设 $a_0,a_1,\cdots,a_n$ 为一组实数，作 $\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)$ 的线性组合，并使其等于零，即

$$a_0\varphi_0(x)+a_1\varphi_1(x)+\cdots+a_n\varphi_n(x)=0$$

分别用$\rho(x)\varphi_0(x),\rho(x)\varphi_1(x),\cdots,\rho(x)\varphi_n(x)$乘上式两边，再在$[a,b]$上积分，根据函数内积的定义，可得方程组

$$\begin{cases}(\varphi_0,\varphi_0)a_0+(\varphi_0,\varphi_1)a_1+\cdots+(\varphi_0,\varphi_n)a_n=0\\(\varphi_1,\varphi_0)a_0+(\varphi_1,\varphi_1)a_1+\cdots+(\varphi_1,\varphi_n)a_n=0\\\cdots\cdots\\(\varphi_n,\varphi_0)a_0+(\varphi_n,\varphi_1)a_1+\cdots+(\varphi_n,\varphi_n)a_n=0\end{cases}$$

由代数学的知识可知，这是关于$a_i(i=0,1,2,\cdots,n)$的齐次方程组，它仅有零解的充分必要条件是其系数行列式（正是线性无关函数的克莱姆行列式）不等于零。

定理证明 若$\{\varphi_i(x)\}$为$[a,b]$上带权$\rho(x)$正交的函数，则依照正交性定义和引理可知，系数矩阵主对角线上的元素均不为零，其他元素全为零，所以$G_n\neq0$，从而$a_i=0$$(i=0,1,2,\cdots,n)$。故$\{\varphi_i(x)\}$为线性无关函数族。

反之由线性无关函数族可选出正交函数族。

二、切比雪夫多项式

1. 切比雪夫多项式的定义

在$[-1,1]$上，形如

$$T_n(x)=\cos(n\cdot\arccos x)\quad(-1\leqslant x\leqslant1)$$

的表达式称为n次切比雪夫多项式。

若记$x=\cos\theta$，则$\theta=\arccos x$，当$x=1$时，$\theta=0$，当$x=-1$时，$\theta=\pi$，于是切比雪夫多项式可以写成

$$T_n(x)=\cos n\theta\quad(0\leqslant\theta\leqslant\pi)$$

2. 切比雪夫多项式的性质

性质1 相邻次数的切比雪夫多项式有以下递推关系：

$$\begin{cases}T_{n+1}(x)=2xT_n(x)-T_{n-1}(x)\quad(n=1,2,\cdots)\\T_0(x)=1,\quad T_1(x)=x\end{cases}$$

证明 由于

$$\cos(n+1)\theta=\cos n\theta\cos\theta-\sin n\theta\sin\theta$$

$$\cos(n-1)\theta=\cos n\theta\cos\theta+\sin n\theta\sin\theta$$

两式相加后移项得

$$\cos(n+1)\theta=2\cos n\theta\cos\theta-\cos(n-1)\theta$$

又

$$T_n(x)=\cos n\theta,\quad T_{n-1}(x)=\cos(n-1)\theta,\quad x=\cos\theta$$

所以

$$T_{n+1}(x)=2xT_n(x)-T_{n-1}(x)$$

由此可得

$$
\begin{aligned}
&T_0(x)=1\\
&T_1(x)=x\\
&T_2(x)=2x^2-1\\
&T_3(x)=4x^3-3x\\
&T_4(x)=8x^4-8x^2+1\\
&T_5(x)=16x^5-20x^3+5x\\
&\cdots\cdots
\end{aligned}
$$

性质2 切比雪夫多项式是关于 x 的 n 次多项式。

证明 根据尤拉公式

$$
\begin{aligned}
T_n(X)=\cos n\theta &=\frac{e^{in\theta}+e^{-in\theta}}{2}\\
&=\frac{(\cos\theta+i\sin\theta)^n+(\cos\theta-i\sin\theta)^n}{2}\\
&=\frac{\left(x+\sqrt{x^2-1}\right)^n+\left(x-\sqrt{x^2-1}\right)^n}{2}
\end{aligned}
$$

分子上两项按牛顿二项式定理展开后，合并，含有 $\sqrt{x^2-1}$ 的奇次方项全部都消为零，所以 $T_n(x)$ 为一个关于 x 的 n 次方多项式。

性质3 切比雪夫多项式 $T_n(x)$ 的最高次项 x^n 的系数为 2^{n-1}。

证明 设 a_n 为 x^n 的系数，则

$$
\begin{aligned}
a_n=\lim_{x\to\infty}\frac{T_n(x)}{x^n}&=\lim_{x\to\infty}\frac{\left(x+\sqrt{x^2-1}\right)^n+\left(x-\sqrt{x^2-1}\right)^n}{2x^n}\\
&=\lim_{x\to\infty}\frac{\left(1+\sqrt{1-\frac{1}{x^2}}\right)^n+\left(1-\sqrt{1-\frac{1}{x^2}}\right)^n}{2}=2^{n-1}
\end{aligned}
$$

性质4 当 $-1\leqslant x\leqslant 1$ 时，$|T_n(x)|\leqslant 1$。

性质5 $T_n(x)$ 在 $[-1,1]$ 中有 n 个不同的实根，且

$$
x_k=\cos\frac{2k-1}{2n}\pi\quad(k=1,2,\cdots,n)
$$

证明 由于 $T_n(x)=\cos(n\cdot\arccos x)=0$ 时

$$
n\cdot\arccos x=k\pi-\frac{\pi}{2}\quad(k=1,2,\cdots,n)
$$

所以

$$
\arccos x=\frac{2k-1}{2n}\pi
$$

$$
x=\cos\frac{2k-1}{2n}\pi
$$

性质6 $T_n(x)$在$[-1,1]$中有$n+1$个点

$$x_k=\cos\left(\frac{k}{n}\pi\right)\quad(k=0,1,2,\cdots,n)$$

轮流取得最大值1和最小值-1。

证明 当$\arccos x=\frac{k}{n}\pi$ $(k=0,1,2,\cdots,n)$时，

$$T_n(x)=\cos(n\cdot\arccos x)=\cos\left(n\frac{k}{n}\pi\right)=\cos(k\pi)=(-1)^k$$

性质7 当n为奇数时，$T_n(x)$为奇函数，当n为偶数时，$T_n(x)$为偶函数。

证明

$$\begin{aligned}T_n(-x)&=\cos[n\cdot\arccos(-x)]\\&=\cos[n\cdot(\pi-\arccos x)]\\&=\cos(n\pi)\cdot\cos(n\cdot\arccos x)\\&=(-1)^n\cdot T_n(x)\end{aligned}$$

性质8 切比雪夫多项式是$[-1,1]$上带权正交的多项式，权函数

$$\rho(x)=(1-x^2)^{-\frac{1}{2}}$$

证明 $(T_m(x),T_n(x))=\int_{-1}^{1}\frac{T_m(x)T_n(x)}{\sqrt{1-x^2}}\mathrm{d}x$

令$x=\cos\theta,\mathrm{d}x=-\sin\theta\mathrm{d}\theta$，当$x=1$时，$\theta=0$；当$x-1$时，$\theta=\pi$。于是

$$(T_m(x),T_n(x))=\int_{\pi}^{0}\frac{\cos(m\theta)\cos(n\theta)}{\sin\theta}(-\sin\theta)\mathrm{d}\theta=\int_0^{\pi}\cos(m\theta)\cos(n\theta)\mathrm{d}\theta$$

当$m\neq n$时，由积化和差分公式得

$$(T_m(x),T_n(x))=\frac{1}{2}\int_0^{\pi}[\cos(m+n)\theta+\cos(m-n)\theta]\mathrm{d}\theta=0$$

当$m=n\neq0$时

$$(T_m(x),T_n(x))=\int_0^{\pi}\cos^2(n\theta)\mathrm{d}\theta=\int_0^{\pi}\frac{1+\cos(2n\theta)}{2}\mathrm{d}\theta=\frac{\pi}{2}$$

当$m=n=0$时

$$(T_m(x),T_n(x))=\int_0^{\pi}\mathrm{d}\theta=\pi$$

故

$$(T_m(x),T_n(x))=\int_{-1}^{1}\frac{T_m(x)T_n(x)}{\sqrt{1-x^2}}\mathrm{d}x=\begin{cases}0 & (m\neq n)\\ \frac{1}{2}\pi & (m=n\neq0)\\ \pi & (m=n=0)\end{cases}$$

由于m,n为任意非负数，则$\{T_n(x)\}$是$[-1,1]$上带权正交的多项式序列，$T_n(x)$为n次正交多项式。

3. 切比雪夫多项式的几何特征

由图 6 - 1 可知切比雪夫多项式有如下的几何特征：

(1) n 次切比雪夫多项式的 n 个实根全部落在[-1,1]上。

(2) $T_n(x)$ 在[-1,1]上的图形被压缩在 $x = \pm 1, y = \pm 1$ 的正方形之中。

(3) $T_n(x)$ 为奇数函数时，过(1,1)与(-1, -1)两点；$T_n(x)$ 为偶数时，过(1,1)与(-1,1)两点。

(4) $T_n(x)$ 在[-1,1]上的极大值、极小值分别落在 $y = 1$ 与 $y = -1$ 上。

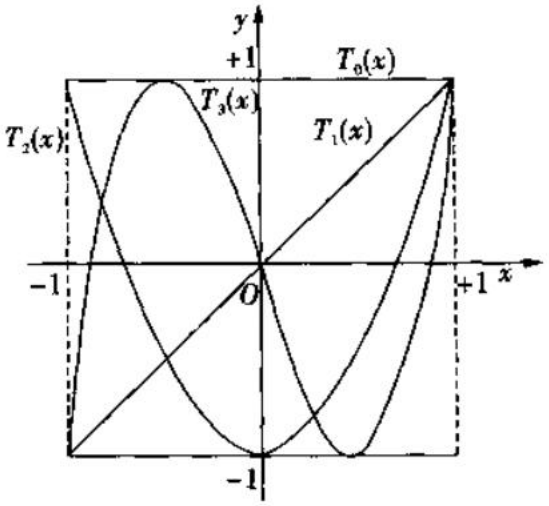


图 6 - 1

第三节　函数的平方逼近——最小二乘法

一、方法的基本思想

1. 插值逼近的固有缺陷

插值法也是函数逼近的一类方法，但是它要求所构造的插值函数 $P_n(x)$ 与要逼近的函数 $f(x)$ 在所给节点处满足

$$f(x_i) = P_n(x_i)$$

由于这种严格的条件限制，给提高插值法的精度带来了较大的困难，具体地说，主要表现在以下两个方面：

(1) 虽然插值法在插值节点处的函数值是精确的，但这是理想化的，实际中节点的函数值通常是以试验方法得到的，这些数据必然会带有这样或那样的误差，这种误差可以说是绝对地进入了插值公式。如果在某些点上的误差较大，则会引起插值函数的严重波动，从而严重影响近似函数的精度。

(2) 一般来说，为了提高插值多项式的精度，可以增加节点的数目，但是随之而来的是插值多项式的次数的提高。所以，无论从高阶差分的角度，还是从对某些函数产生的严重波动现象都可以知道，利用高次插值多项式时，不仅计算烦琐，而且逼近效果也未必

理想。这也就是实际中往往采用低次分段插值的道理。

鉴于以上原因，产生了另一类不依赖于插值节点处函数值精确的条件的逼近方法。下面讨论的函数平方逼近——最小二乘法就是其中之一。

2. 最小二乘法

现在这样来考虑问题：对于表格函数给出的一组实验数据$(x_i,y_i)(i=1,2,\cdots,m)$，不要求满足在已知点处函数值精确，而是从一个函数类$\{\varphi_j(x)\}$中选取一个函数$\varphi(x)$，使它能够成为$f(x)$的最佳逼近函数，即

$$f(x)\approx\varphi(x)$$

这里涉及到一个最好的评价标准问题，即希望对于给定的一批点$(x_i,y_i)(i=1,2,\cdots,m)$，在每一个点上函数$\varphi(x)$对于$f(x)$的偏差

$$\delta_i=f(x_i)-\varphi(x_i)\quad(i=1,2,\cdots,m)$$

都相对较小，面从总体上这些偏差的和能够达到最小。于是，问题就转化为以总偏差作为目标函数来研究它的极小值问题。

然而，很明显不能把总偏差作为数学模型，即不能确定目标函数为

$$D=\sum_{i=1}^{m}\delta_i$$

这里由于δ_i可正可负，尽管D可能很小，但$|\delta_i|$的值可能很大。所以D不能作为评价函数$\varphi(x)$好坏的标准。为了避免正负偏差可能抵消的因素，自然就想到用偏差的平方和

$$E=\sum_{i=1}^{m}\delta_i^2=\sum_{i=1}^{m}\left[\varphi(x_i)-f(x_i)\right]^2$$

作为目标函数来讨论偏差是否合适。

3. 最小二乘法的数学描述

对于给定的一组试验数据$(x_i,y_i)(i=1,2,\cdots,m)$，要求从函数类

$$\Phi=\{\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)\}$$

中找到一个函数

$$\varphi(x)=a_0\varphi_0(x)+a_1\varphi_1(x)+\cdots+a_n\varphi_n(x)\quad(n<m)$$

使$\varphi(x)$满足在权函数$\rho(x)$下，差的平方和取极小值，即

$$F(x)=\sum_{i=1}^{m}\rho(x)\cdot\left[\varphi_i(x)-f(x_i)\right]^2$$

使$F(x)$取最小值，而权函数$\rho(x)$描述了x_i点上函数值的重要程度或同样数据出现的重复次数等。

二、最小二乘法的一般解法

1. 逼近函数类的选择

从上一段的描述中可知，平方逼近问题主要是从一个函数类中选择一个最好的逼近函数的问题，所以，选择函数的问题就成为解最小二乘法的首要问题，也是关键的问题，

选的不好是绝对达不到逼近要求的，这不单纯是一个数学问题，更重要的是专业知识和实际经验。常用的方法是用给定的点(x_i,y_i)描出草图，以此来猜测逼近函数的类型。通常是选择函数类中某些函数的线性组合，随后结合例题来说明这种选择方法，实在不行时，重新进行选择。

2. 最小二乘法的一般解法

下边的讨论仅就$\rho(x)=1$的情形进行。对于要求满足

$$F(x)=\min_{\varphi(x)\in\Phi}\sum_{i=1}^{m}[\varphi(x_i)-f(x_i)]^2$$

的最小二乘法问题，实际上是确定函数(目标函数)

$$E(a_0,a_1,\cdots,a_n)=\sum_{i=1}^{m}[\varphi(x_i)-f(x_i)]^2$$

的极小点$(a_0^*,a_1^*,\cdots,a_n^*)$的问题。

这是一个多元函数的极值问题，可以通过解方程组来解决。

由

$$\frac{\partial E}{\partial a_k}=0\quad(k=0,1,2,\cdots,n)$$

即

$$\frac{\partial E}{\partial a_k}\Big\{\sum_{i=1}^{m}[a_0\varphi_0(x_i)+a_1\varphi_1(x_i)+\cdots+a_n\varphi_n(x_i)-f(x_i)]^2\Big\}=0$$

亦即

$$\sum_{i=1}^{m}\varphi_k(x_i)[a_0\varphi_0(x_i)+a_1\varphi_1(x_i)+\cdots+a_n\varphi_n(x_i)-f(x_i)]=0$$

所求出的a_k即为最小点a_k^* $(k=0,1,2,\cdots,n)$，当使用内积的记号时，

$$\begin{cases}(\varphi_k,\varphi_j)=\sum_{i=1}^{m}\varphi_k(x_i)\varphi_j(x_i)\\ (\varphi_k,f)=\sum_{i=1}^{m}\varphi_k(x_i)f(x_i)\end{cases}\qquad(k,j=0,1,\cdots,n)$$

于是方程组可写为

$$\begin{cases}a_0(\varphi_0,\varphi_0)+a_1(\varphi_0,\varphi_1)+\cdots+a_n(\varphi_0,\varphi_n)=(\varphi_0,f)\\ a_0(\varphi_1,\varphi_0)+a_1(\varphi_1,\varphi_1)+\cdots+a_n(\varphi_1,\varphi_n)=(\varphi_1,f)\\ \cdots\cdots\\ a_0(\varphi_n,\varphi_0)+a_1(\varphi_n,\varphi_1)+\cdots+a_n(\varphi_n,\varphi_n)=(\varphi_n,f)\end{cases}$$

其矩阵形式为

$$\begin{bmatrix} (\varphi_0,\varphi_0) & (\varphi_0,\varphi_1) & \cdots & \varphi_0,\varphi_n \\ (\varphi_1,\varphi_0) & (\varphi_1,\varphi_1) & \cdots & (\varphi_1,\varphi_n) \\ \vdots & \vdots & & \vdots \\ (\varphi_n,\varphi_0) & (\varphi_n,\varphi_1) & \cdots & (\varphi_n,\varphi_n) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} (\varphi_0,f) \\ (\varphi_1,f) \\ \vdots \\ (\varphi_n,f) \end{bmatrix}$$

该线性方程组的行列式是函数 $\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)$ 的克莱姆行列式 G_n。从前边的讨论可知，只要这些函数线性无关，方程组就有唯一的一组解：

$$a_0 = a_0^*,\quad a_1 = a_1^*,\quad \cdots,\quad a_n = a_n^*$$

于是，最佳平方逼近函数为

$$\varphi(x) = a_0^*\varphi_0(x) + a_1^*\varphi_1(x) + \cdots + a_n^*\varphi_n(x)$$

通常，称上面的方程组为法方程组，或称为正则方程组，称其系数矩阵为法矩阵或正则矩阵。

上述求解最小二乘法的过程可概括为以下两步：

(1)根据已知点 $(x_i,y_i)(i=1,2,\cdots,m)$，在各种函数中适当地选择由线性无关的函数 $\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)$ 构成的函数类，再用这些线性无关的函数的某个线性组合 $\varphi(x)=a_0\varphi_0(x)+a_1\varphi_1(x)+\cdots+a_n\varphi_n(x)$ 作为 $f(x)$ 的最佳平方逼近函数。

(2)用满足 $\frac{\partial E}{\partial a_k}=0(k=0,1,2,\cdots,n)$ 的正则方程组来确定最佳点 $(a_0^*,a_1^*,\cdots,a_n^*)$。

三、正则矩阵的形成与计算

从正则矩阵的形式可以看出，它的元素是由线性无关的函数组 $\{\varphi_k(x)\}$ 中的函数在给定点 x_i 上两两作成的内积，而由内积的可交换性可知，正则矩阵是对称矩阵。依此两点，可以给出形成正则矩阵的简便方法，这个方法容易在电子计算机上实现。具体作法如下：

由线性无关函数组 $\{\varphi_k(x)\}$ 作成下述矩阵：

$$C = \begin{bmatrix} \varphi_0(x_0) & \varphi_0(x_1) & \cdots & \varphi_0(x_m) \\ \varphi_1(x_0) & \varphi_1(x_1) & \cdots & \varphi_1(x_m) \\ \vdots & \vdots & & \vdots \\ \varphi_n(x_0) & \varphi_n(x_1) & \cdots & \varphi_n(x_m) \end{bmatrix}$$

称矩阵 C 为中间矩阵，则 C^T 为

$$C^T = \begin{bmatrix} \varphi_0(x_0) & \varphi_1(x_0) & \cdots & \varphi_n(x_0) \\ \varphi_0(x_1) & \varphi_1(x_1) & \cdots & \varphi_1(x_n) \\ \vdots & \vdots & & \vdots \\ \varphi_0(x_m) & \varphi_1(x_m) & \cdots & \varphi_n(x_m) \end{bmatrix}$$

于是，正则矩阵 A 可由下式确定

$$A = C \cdot C^T$$

因 C 为 $(n+1)\times m$ 阶矩阵，C^T 为 $m\times(n+1)$ 阶矩阵，则 A 是一个 $(n+1)\times(n+1)$ 阶的方阵。不难验证，A 的任一个元素

$$(\varphi_k,\varphi_j)=\sum_{i=1}^{m}\varphi_k(x_i)\varphi_j(x_i)\quad(k,j=0,1,2,\cdots,n)$$

而正则方程组中的右端项的计算可用

$$(\varphi_k,f)=\sum_{i=1}^{m}\varphi_k(x_i)f(x_i)\quad(k=0,1,2,\cdots,n)$$

来完成。

四、最小二乘法的计算步骤

由上所述，最小二乘法的计算步骤可归纳为：

(1)输入 $(x_i,y_i)(i=1,2,\cdots,m)$，由 (x_i,y_i) 选出

$$\varphi(x)=a_0\varphi_0(x)+a_1\varphi_1(x)+\cdots+a_n\varphi_n(x)$$

(2)按线性无关函数组形成中间矩阵 C，且 $B=C^T$；

(3)形成正则矩阵及右端项

$$A=CB,\quad G=(\varphi_k,f)$$

即

$$a_{kj}=(\varphi_k,\varphi_j)=\sum_{i=1}^{m}\varphi_k(x_i)\varphi_j(x_i)\quad(k,j=0,1,2,\cdots,n)$$

$$g_k=(\varphi_k,f)=\sum_{i=1}^{m}\varphi_k(x_i)f(x_i)\quad(k=0,1,2,\cdots,n)$$

(4)解正则方程组 A，确定 $\{a_k\}$；

(5)输出 $\{a_k\}$，得

$$\varphi(x)=a_0\varphi_0(x)+a_1\varphi_1(x)+\cdots+a_n\varphi_n(x)$$

第四节　最小二乘法多项式的逼近

由于多项式函数具有计算简单以及其他一些好的性质，所以实际中常选用多项式作为最佳平方逼近函数。有了上节的知识作为基础，把它具体应用在多项式即可。

一、正则方程组的建立

现选取线性无关函数组为

$$\{\varphi_k(x)\}=\{\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)\}=\{1,x,x^2,\cdots,x^n\}$$

于是，正则方程组即为

$$\begin{bmatrix} m & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^n \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \cdots & \sum x_i^{n+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum x_i^n & \sum x_i^{n+1} & \sum x_i^{n+2} & \cdots & \sum x_i^{2n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum f(x_i) \\ \sum x_i f(x_i) \\ \vdots \\ \sum x_i^n f(x_i) \end{bmatrix}$$

其中，$\sum$ 为 $\sum_{i=1}^{m}$；m 为表格函数中已知点个数，即试验点数。

二、正则矩阵的生成方法

由于正则矩阵是对称的，因此可以给出多种生成方法。所谓生成一个矩阵是指计算出该矩阵的各个元素。现结合用多项式作平方逼近，且线性无关组取$\{1,x,x^2,\cdots,x^n\}$的情形给出3种生成正则矩阵的方法。

1. 中间矩阵法

按照线性无关组$\{1,x,x^2,\cdots,x^n\}$构造中间矩阵

$$C = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & & \vdots \\ x_1^n & x_2^n & \cdots & x_m^n \end{bmatrix}$$

则正则矩阵为$A = C \cdot C^{\mathrm{T}} = C \cdot B$，这里$B = C^{\mathrm{T}}$。这种方法要用到一个$(n+1)\times(n+1)$阶的方阵$A$以及两个$(n+1)\times m$的阶的矩阵$C$和$B$。相应地在计算机程序中就要定义三个二维数组，要作一次矩阵转置和矩阵的乘法。

2. 直接生成法

直接用线性无关组$\{1,x,x^2,\cdots,x^n\}$中各元素在给定点上的值算正则矩阵A的各个元素

$$(\varphi_k,\varphi_j) = \sum_{i=1}^{m} \varphi_k(x_i)\varphi_j(x_i) = \sum_{i=1}^{m} x_i^{k-1}x_i^{j-1} = \sum_{i=1}^{m} x_i^{k+j-2} \quad (k,j = 1,2,\cdots,n+1)$$

在计算机程序中，可用一个三重循环结构来计算正则矩阵A的每一个元素，即

$$A(k,j) = (\varphi_k,\varphi_j) = \sum_{i=1}^{m} x_i^{k+j-2}$$

使用直接法，只需要定义一个二维数组$A(n+1,n+1)$，但同样要作一个矩阵乘法的运算。显然，程序的编制比较简单，只要一个三重循环就可以了。

3. 反对角线赋值法

在用多项式作平方逼近时，若选取了线性无关组$\{1,x,x^2,\cdots,x^n\}$，则正则矩阵不仅对称，而且沿反对角线上的元素全相等。

利用这一特点，只要计算正则矩阵A的第一行上的$n+1$个元素和最后一个行的n个元素（因为第一行上的最后一个元素和最后一行的第一个元素相等），共计$2n+1$个元

素就可以了。每当算出一个元素时，沿其反对角线对其他元素赋以同样的数值即可。

第一行和最后一行（或最后一列）的 $2n+1$ 个元素可表示为

$$S_k = \sum_{i=1}^{m} x_i^k \quad (k = 0,1,2,\cdots,2n)$$

显然

$$S_0 = \sum_{i=1}^{m} x_i^0 = m, \quad S_1 = \sum_{i=1}^{m} x_i, \quad \cdots, \quad S_{2n} = \sum_{i=1}^{m} x_i^{2n}$$

这样，矩阵 A 中各元素值为

$$a_{kj} = a_{jk} = S_{k+j-2} = \sum_{i=1}^{m} x_i^{k+j-2} \quad (k,j = 1,2,\cdots,n+1)$$

很明显，这个算法优于前面两个算法，存贮量与第二种方法相当，但运算次数降低了一个数量级，因为它避免了矩阵的乘法。当矩阵的阶数较高时，这种方法在时间与存贮空间上都能节省。

三、最小二乘法多项式逼近的计算步骤

通常，当进行多项式逼近时，参与拟合的点对数 m 应大于所拟合多项式的次数 n，若 $m<n$，则拟合效果不良，故求定多项式型经验公式时，应测定足够的实验数据。

对于同一个表格函数 $(x_i,y_i)(i=1,2,\cdots,m)$，可用不同项的多项式来拟合，但一般当 $n\geqslant 7$ 时，易产生病态的正则矩阵，因此，可在 $2\leqslant n<7$ 的范围内选择 n，以期达到要求的精度。通常采用总相对误差作为 n 值选择的判据，即

$$\frac{\sum_{i=1}^{m} |\varphi(x_i) - f(x_i)|}{\left|\sum_{i=1}^{m} \varphi(x_i)\right|} \leqslant \varepsilon$$

最后应指出的是，最小二乘法多项式逼近与其他最小二乘拟合法一样，事后要进行显著性检验。只有结果显著时，才能用多项式这类函数去逼近所述的离散函数，否则应采用其他适宜的函数形式去逼近。

综上所述，最小二乘法多项式逼近的计算步骤为：

(1) 让 $n=2$，输入 (x_i,y_i) 和 ε；

(2) 计算 A 的各元素及右端自由项

$$a_{kj} = a_{jk} = S_{k+j-2} = \sum_{i=1}^{m} x_i^{k+j-2} \quad (k,j = 1,2,\cdots,n+1),$$

$$a_{k\,n+2} = \sum_{i=1}^{m} x_i^{k-1} \cdot f(x_i);$$

(3) 利用 LDL^T 分解法求 $a_0,a_1,\cdots,a_n$；

(4) 计算总相对误差

$$d=\frac{\sum_{i=1}^{m}|f(x_i)-\varphi(x_i)|}{\left|\sum_{i=1}^{m}\varphi(x_i)\right|};$$

(5)若 $d<\varepsilon$,输出 $\varphi(x)=a_0+a_1x+\cdots+a_nx^n(x)$,停机;

(6)若 $n<7$,则令 $n=n+1$ 回第(2)步;否则,打印奇异标志,并输出 $a_0,a_1,\cdots,a_n$ 及总相对误差 d,停机。

第五节　经验公式的使用及非线性函数的线性化

本节结合例题来介绍在实际中使用最小二乘法时处理问题的常用方法。主要说明如何确定拟合函数的类型、常用的经验公式以及非线性函数的线性方法等问题。

一、如何确定逼近函数的类型

通常是把给定的点在坐标纸上按趋势绘成一个草图,再由图形的形状估测出函数的类型,以作为进一步精确化的基础。现举例如下。

例 6-3　已知一批实验数据如下表所示,试用最小二乘法求其拟合曲线的表达式。

序号 i	1	2	3	4	5	6	7	8	9	10	11
x_i	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
y_i	8.50	6.70	4.90	3.70	3.00	2.70	3.10	3.80	4.90	6.65	9.00

解　由给定的点作草图 6-2,由图 6-2 可见,这大体上是一个抛物线。则可选取二次多项式为逼近函数,其函数的线性无关组可选 $\{1,x,x^2\}$,且这时 $m=11,n=2$,则正则方程组为

$$\begin{bmatrix} 11 & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix}\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}=\begin{bmatrix} \sum y_i \\ \sum x_iy_i \\ \sum x_i^2y_i \end{bmatrix}$$

把所给的 (x_i,y_i) 代入,再展开方程组得

$$\begin{cases} 11a_0+27.5a_1+96.25a_2=56.92 \\ 27.5a_0+96.25a_1+378.125a_2=143.675 \\ 96.25a_0+378.125a_1+1\,583.313a_2=556.288 \end{cases}$$

解得

$$a_0=8.659,\quad a_1=-4.752,\quad a_2=0.960$$

所以

$$y = 8.659 - 4.752x + 0.960x^2$$

总相对误差为

$$d = \frac{\sum_{i=1}^{11} |y_i - y(x_i)|}{\sum_{i=1}^{11} |y(x_i)|} = \frac{0.809}{56.96}$$

$$= 0.014\ 2$$

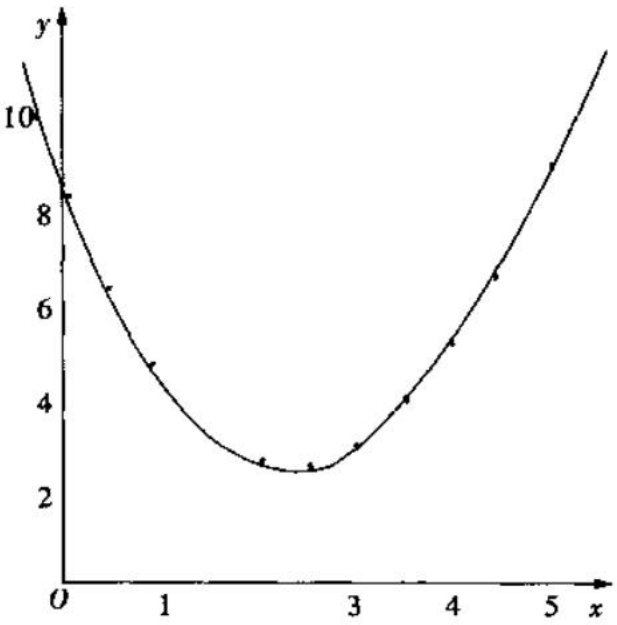


图 6-2

二、常用的几种经验公式及线性化方法

前述的寻求最佳平方逼近函数是通过确定一个线性无关组的线性组合作为逼近函数，这仅是构造逼近函数的一种方法，对于多项式逼近的情形，这种方法特别方便。但是，从最小二乘法的基本原理上说，这并不是唯一的方法。

事实上，人们在使用最佳平方逼近的长期实践中发现，许多函数表达式，即所谓经验公式是特别常用的逼近函数。这些经验公式中的未定常数也应由给出的实验点来确定。然而，当用最小二乘法构造出确定这些特定常数的数学模型时，将会发现，往往要解一个非线性的方程组。这就给计算带来了许多不便。为些，首先要对其进行线性化，现给出几种常用的经验公式线性化的方法。

(1) 经验公式：$\varphi(t) = \dfrac{1}{a + bt}$

线性化方法：令 $y = \dfrac{1}{\varphi(t)}$，$x = t$，则 $y = a + bx$。

(2)经验公式:$\varphi(t)=\dfrac{t}{at+b}$

线性化方法:令 $y=\dfrac{1}{\varphi(t)}$,$x=\dfrac{1}{t}$,则 $y=a+bx$。

(3)经验公式:$\varphi(t)=q\cdot t^{p}$

线性化方法:两边取对数得 $\ln\varphi(t)=\ln q+p\cdot\ln t$

令 $y=\ln\varphi(t)$,$a=\ln q$,$b=p$,$x=\ln t$,则 $y=a+bx$。

(4)经验公式:$\varphi(t)=qe^{pt}$

线性化方法:两边取对数得 $\ln\varphi(t)=\ln q+pt$

令 $y=\ln\varphi(t)$,$a=\ln q$,$b=p$,$x=t$,则 $y=a+bx$。

应当注意的是,经验公式经线性化之后,即成为一次多项式,可按多项式逼近处理,并在线性化之后相应地算出新函数及自变量的值,以使得能确定系数 a、b,然后回代确定原参数,得逼近函数,最后进行显著性检验。

例 6-4 在间歇反应器中进行动力学实验,测得其生成物的浓度随时间变化关系为:

t/min	1	2	3	4	5	6	7	8
$c\times10^{3}$/(mol·L^{-1})	4.00	6.00	8.00	8.80	9.22	9.50	9.70	9.86
t/min	9	10	11	12	13	14	15	16
$c\times10^{3}$/(mol·L^{-1})	10.00	10.20	10.32	10.42	10.50	10.55	10.58	10.60

试用最小二乘法确定 $c(t)$ 的函数关系。

解 首先由已知点标绘 $c-t$ 曲线(见图 6-3),由标绘曲线可知,$c-t$ 数据可表示为

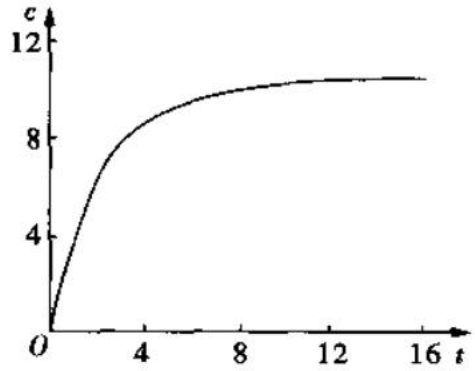


图 6-3

$$c=q\exp(p/t)$$

即

$$\ln c = \ln q + \frac{p}{t}$$

令

$$y = \ln c, \quad x = \frac{1}{t}, \quad a = \ln q, \quad b = P$$

则

$$y = a + bx$$

于是，对线性化以后的函数即可进行线性逼近，即选取线性无关组 $\{1, x\}$。此时，$m = 16, n = 1$，其相应的正则方程组为

$$\begin{bmatrix} 16 & 3.38 \\ 3.38 & 1.58 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -75.33 \\ -16.85 \end{bmatrix}$$

解得

$$a = -4.4823, \quad b = -1.0689$$

于是

$$q = 1.13 \times 10^{-2}, \quad p = -1.0689$$

则

$$c = 1.13 \times 10^{-2} \exp\left(-\frac{1.0689}{t}\right)$$

$$d = \frac{\sum_{i=1}^{16} |c_i - c(t_i)|}{\sum_{i=1}^{16} |c(t_i)|} = \frac{0.0016}{0.14795} = 0.0108$$

计算出相关指数 $R^2 = 0.99$，拟合良好。

第六节　利用切比雪夫多项式的平方逼近

由前几节的讨论，不仅熟悉了平方逼近的原理，而且也学会了最小二乘法的解法。但是实践证明，当多项式的次数 $n \geq 7$ 时，所得到的正则方程组往往是病态的，也即正则矩阵是病态的。病态方程组的数值解会产生很大的误差，是很不可靠的。

为了避免这种情况，产生了许多求最小二乘法解的新方法。这里仅讨论利用切比雪夫多项式求最小二乘法解的方法及其基本原理。

一、方法的基本原理

切比雪夫多项式有许多好的性质，如它在 $[-1, 1]$ 上是带权 $\rho(x) = (1 - x^2)^{\frac{1}{2}}$ 的正交多项式族，于是就取各次切比雪夫多项式作为最小二乘法的线性无关函数组，即选

取$\{T_0(x),T_1(x),\cdots,T_n(x)\}$,用它们的线性组合构造的最佳平方逼近函数为

$$y(x)=a_0T_0(x)+a_1T_1(x)+\cdots+a_nT_n(x)$$

但是,为了确定这些待定的$a_k(k=0,1,\cdots,n)$,首先应解决以下两个问题:

(1)切比雪夫多项式是在$[-1,1]$上的正交多项式,如何把给定在任意区间$[a,b]$上的实验点x_i映射在$[-1,1]$上,成为t_i,且这种变换可以保证相应函数值y_i不变。

(2)应能使这样产生的确定a_k的正则矩阵摆脱病态,从而便于求出a_k。

第一个问题可通过线性变换

$$x_i=\frac{a+b}{2}+\frac{b-a}{2}t_i \quad (i=1,2,\cdots,n)$$

即可使$x\in[a,b]$映射到$[-1,1]$上。

要解决第二个问题,需借助于以下定理作为理论基础。

定理 设$x_i(i=1,2,\cdots,n)$为$n+1$次切比雪夫多项式的零点,即$x_i=\cos\frac{2i-1}{2(n+1)}\pi$,则对于任何次数不高于$n$次的切比雪夫多项式$T_j(x)$与$T_k(x)$(这里,$j,k\leqslant n$),有以下性质:

$$(T_j(x),T_k(x))=\sum_{i=1}^{n+1}T_j(x_i)T_k(x_i)=\begin{cases}0 & (j\neq k)\\ \frac{n+1}{2} & (j=k\neq 0)\\ n+1 & (j=k=0)\end{cases}$$

证明(1)先考察$j=k=0$的情形,这时

$$T_j(x_i)=T_k(x_i)=T_0(x_i)=1 \quad (i=1,2,\cdots,n+1)$$

则

$$(T_j,T_k)=\sum_{i=1}^{n+1}=n+1$$

(2)若j,k中至少有一个不为零时,按照切比雪夫多项式的定义以及x_i是$T_{n+1}(x)$的零点,再利用三角函数的积化和差公式可推得

$$\begin{aligned}(T_j,T_k)&=\sum_{i=1}^{n+1}T_j(x_i)T_k(x_i)=\sum_{i=1}^{n+1}[\cos(j\arccos x_i)\cdot\cos(k\arccos x_i)]\\&=\sum_{i=1}^{n+1}\left\{\cos\left[j\frac{2i-1}{2(n+1)}\pi\right]\cdot\cos\left[k\cdot\frac{2i-1}{2(n+1)}\pi\right]\right\}\\&=\frac{1}{2}\left\{\sum_{i=1}^{n+1}\cos\left[\frac{(2i-1)(j+k)}{2(n+1)}\pi\right]+\sum_{i=1}^{n+1}\cos\left[\frac{(2i-1)(j-k)}{2(n+1)}\pi\right]\right\}\end{aligned}$$

令

$$S_1=\sum_{i=1}^{n+1}\cos\left[\frac{(2i-1)(j+k)}{2(n+1)}\pi\right],\quad S_2=\sum_{i=1}^{n+1}\cos\left[\frac{(2i-1)(j-k)}{2(n+1)}\pi\right]$$

则

$$(T_j, T_k) = \frac{1}{2}(S_1 + S_2)$$

又

$$\cos\alpha = \frac{\sin(\alpha+\beta) - \sin(\alpha-\beta)}{2\sin\beta}$$

记

$$\alpha = \frac{(2i-1)(j+k)}{2(n+1)}\pi, \quad \beta = \frac{(j+k)}{2(n+1)}\pi$$

于是

$$S_1 = \sum_{i=1}^{n+1} \frac{\sin\left[\frac{(j+k)}{n+1}i\pi\right] - \sin\left[\frac{(i-1)(j+k)}{n+1}\pi\right]}{2\sin\left[\frac{(j+k)}{2(n+1)}\pi\right]}$$

$$= -\frac{1}{2\sin\left[\frac{(j+k)}{2(n+1)}\pi\right]} \sum_{i=1}^{n+1} \left\{\sin\left[\frac{(j+k)}{n+1}i\pi\right] - \sin\left[\frac{(j+k)(i-1)}{n+1}\pi\right]\right\}$$

而

$$\sum_{i=1}^{n+1} \left\{\sin\left[\frac{(j+k)i}{n+1}\pi\right] - \sin\left[\frac{(j+k)(i-1)}{n+1}\pi\right]\right\}$$

$$= \sin\left[\frac{j+k}{n+1}\pi\right] - 0 + \sin\left[\frac{2(j+k)}{n+1}\pi\right] - \sin\left[\frac{j+k}{n+1}\pi\right] + \sin\left[\frac{3(j+k)}{n+1}\pi\right] -$$

$$\sin\left[\frac{2(j+k)}{n+1}\pi\right] + \cdots + \sin[(j+k)\pi] - \sin\left[\frac{n(j+k)}{n+1}\pi\right]$$

$$= \sin[(j+k)\pi]$$

所以

$$S_1 = \frac{\sin[(j+k)\pi]}{2\sin\left[\frac{j+k}{2(n+1)}\pi\right]}$$

显然，$\sin(j+k)\pi = 0$，由于 $j, k \leqslant n$ 且至少有一个不为零，则分母不为零。于是 $S_1 = 0$。

若又记

$$\alpha = \frac{(2i-1)(j+k)}{2(n+1)}\pi, \quad \beta = \frac{j-k}{2(n+1)}\pi$$

则

$$S_2 = \sum_{i=1}^{n+1} \frac{\sin\left[\frac{j-k}{n+1}i\pi\right] - \sin\left[\frac{(j-k)(i-1)}{n+1}\pi\right]}{2\sin\left[\frac{(j-k)}{2(n+1)}\pi\right]}$$

$$= -\frac{1}{2\sin\left[\frac{j-k}{2(n+1)}\pi\right]} \sum_{i=1}^{n+1} \left\{\sin\left[\frac{j-k}{n+1}i\pi\right] - \sin\left[\frac{(j-k)(i-1)}{n+1}\pi\right]\right\}$$

$$= \frac{\sin[(j-k)\pi]}{2\sin\left[\frac{j-k}{2(n+1)}\pi\right]}$$

当$j \neq k$时

$$\sin[(j-k)\pi] = 0, \quad \sin\left[\frac{j-k}{2(n+1)}\pi\right] \neq 0$$

于是$S_2 = 0$。

当$j = k$时

$$S_2 = \lim_{j \to k} \frac{\sin[(j-k)\pi]}{2\sin\left[\frac{(j-k)}{2(n+1)}\pi\right]} = n+1$$

于是

$$S_2 = \begin{cases} 0 & (j \neq k) \\ n+1 & (j = k \neq 0) \end{cases}$$

这样

$$(T_j, T_k) = \frac{1}{2}(S_1 + S_2) = \begin{cases} n+1 & (j = k = 0) \\ \frac{1}{2}(n+1) & (j = k \neq 0) \\ 0 & (j \neq k) \end{cases}$$

二、最小二乘解的确定

1. 直接结果

上述定理表明，给定的曲线拟合区间确实在$[-1,1]$上，且实验数据也在m次切比雪夫多项式$T_m(x)$的零点$x_i(i=1,2,\cdots,m)$上给出，其中，要求$m > n$，即$m \geqslant n+1$。

于是，由线性无关组$\{T_k(x)\}(k=1,2,\cdots,n)$产生的确定最小二乘解的正则方程组为

$$\begin{bmatrix} (T_0,T_0) & & & \\ & (T_1,T_1) & & \\ & & \ddots & \\ & & & (T_n,T_n) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} (T_0,f) \\ (T_1,f) \\ \vdots \\ (T_n,f) \end{bmatrix}$$

这是一个由对角矩阵组成的方程组，而且由上述定理保证了主对角线上的元素$(T_k,T_k) \neq 0$，这就使正则矩阵摆脱了病态，并且要确定的a_k的解为

$$a_k = \frac{(T_k,f)}{(T_k,T_k)} = \frac{\sum_{i=1}^{m} T_k(x_i) f(x_i)}{\sum_{i=1}^{m} T_k(x_i) T_k(x_i)} \qquad (k = 1,2,\cdots,n)$$

则

$$a_k = \begin{cases} \dfrac{1}{m}\sum_{i=1}^{m} f(x_i) & (k=0) \\ \dfrac{2}{m}\sum_{i=1}^{m} T_k(x_i) f(x_i) & (k=1,2,\cdots,n) \end{cases}$$

2. 一般求解过程

现在来讨论当给定的拟合区间不是$[-1,1]$，而是$[a,b]$，同时给出的表格函数也未必是切比雪夫多项式零点的情形。解决这种一般问题的方法是：

首先，把给定的点$z_i \in [a,b]$变换为$t_i \in [-1,1]$$(i=1,2,\cdots,m)$，即

$$z_i = \frac{a+b}{2} + \frac{b-a}{2} t_i$$

其次，根据函数表(t_i, y_i)构造一个插值多项式，并求出$n+1$次切比雪夫多项式$T_{n+1}(x)$的零点

$$x_j = \cos\left[\frac{2j-1}{2(n+1)}\pi\right] \quad (j=1,2,\cdots,n+1)$$

以及对应x_i处的函数值y_j。

最后，由这些点(x_i, y_i)和线性无关组$\{T_k(x)\}$$(k=1,2,\cdots,n)$产生确定最小二乘法的正则方程组求解$a_k(k=1,2,\cdots,n)$，得到在区间$[-1,1]$上的多项式

$$\Phi_n(t) = a_0 T_0(t) + a_1 T_1(t) + \cdots + a_n T_n(t)$$

再把区间变回到$[a,b]$上，则上式成为

$$y_n(z) = \Phi_n\left(\frac{2z-(a+b)}{b-a}\right) = A_0 + A_1 z + A_2 z^2 + \cdots + A_n z^n$$

下边结合例子来说明实现这个一般过程的具体方法。

例6-5 已知一组实验数据$(z_i, y_i)$$(i=1,2,\cdots,7)$，如下表所示。试用切比雪夫多项式求其二次多项式拟合曲线。

z_i	1.000 0	1.500 0	2.000 0	2.500 0	3.000 0	3.500 0	4.000 0
y_i	4.900 0	3.700 0	3.000 0	2.700 0	3.100 0	3.800 0	4.900 0

解 现将求解过程分为以下几步：

(1)作线性变换$z_i = \frac{a+b}{2} + \frac{b-a}{2} t_i$，则

$$t_i = \frac{2z_i - (a+b)}{b-a} \quad (i=1,2,\cdots,7)$$

把$z \in [a,b]$映射到$t \in [-1,1]$，这里$a=1, b=4$，于是

$$t_i = \frac{1}{3}(2z_i - 5) \quad (i=1,2,\cdots,7)$$

计算结果列入表6－1的第三行中。

表6－1

z_i	1.000 0	1.500 0	2.000 0	2.500 0	3.000 0	3.500 0	4.000 0
y_i	4.900 0	3.700 0	3.000 0	2.700 0	3.100 0	3.800 0	4.900 0
t_i	－1.000 0	－0.666 7	－0.333 3	0.000 0	0.333 3	0.666 7	1.000 0
x_j	0.866 0	0.000 0	－0.866 0				
y_j	4.457 8	2.700 0	4.417 7				

(2)计算 $T_{n+1}(x)$ 的零点

$$x_j=\cos\left[\frac{(2j-1)}{2(n+1)}\pi\right] \quad (j=1,2,\cdots,n+1)$$

一般是先确定拟合多项式的次数 n，然后再确定 m，通常取 $m=n+1$。

这里 $n=2$，故按照定理的要求至少应算出三次切比雪夫多项式 $T_3(x)$ 的各个零点，即

$$x_1=\cos\frac{\pi}{6}=0.866\ 0$$

$$x_2=\cos\frac{\pi}{2}=0.000\ 0$$

$$x_3=\cos\frac{5\pi}{6}=-0.866\ 0$$

这些数据已列入表6－1第四行中。

(3)根据 $(t_i,y_i)(i=1,2,\cdots,7)$ 构造拉格朗日插值公式，计算在 $x_j(j=1,2,3)$ 处的函数值 y_j。已将用分段性插法计算的 y_j 值列入表6－1的第五行中。

(4)计算 $a_k(k=0,1,\cdots,n)$，这里 $k=0,1,2$，

$$a_k=\begin{cases}\dfrac{1}{m}\displaystyle\sum_{j=1}^{m}y_j & (k=0)\\ \dfrac{2}{m}\displaystyle\sum_{j=1}^{m}T_k(x_j)y_j & (k=1,2,\cdots,n)\end{cases}$$

这里 $n=2,m=2+1=3$，则

$$a_0=\frac{1}{3}(y_1+y_2+y_3)=\frac{1}{3}(4.457\ 8+2.700\ 0+4.417\ 7)=3.858\ 5$$

$$a_1=\frac{2}{3}[y_1T_1(x_1)+y_2T_1(x_2)+y_3T_1(x_3)]$$

$$=\frac{2}{3}(4.457\ 8\times0.866\ 0\times2.700\ 0\times0.000\ 0-4.417\ 7\times0.866\ 0)$$

$$=0.023\ 2$$

$$a_2=\frac{2}{3}[y_1T_2(x_1)+y_2T_2(x_2)+y_3T_2(x_3)]$$

$$=\frac{2}{3}(4.457\ 8\times0.499\ 9\times2.700\ 0\times(-1)+4.417\ 7\times0.499\ 9)$$

$$=1.158\ 0$$

(5)写出在区间$[-1,1]$上的切比雪夫拟合多项式

$$\Phi_n(t)=a_0T_0(t)+a_1T_1(t)+\cdots+a_nT_n(t)$$

这里$n=2$,则

$$\Phi_2(t)=a_0T_0(t)+a_1T_1(t)+a_2T_2(t)$$

$$\Phi_2(t)=3.858\ 5+0.023\ 2t+1.158\ 0(2t^2-1)$$

即

$$\Phi_2(t)=2.700\ 5+0.023\ 2t+2.316\ 0t^2$$

(6)利用逆变换映射回原来的区间$[a,b]$。由于

$$t=\frac{2z-(a+b)}{b-a},\quad y_n(z)=A_0+A_1z+\cdots+A_nz^n$$

当区间为$[1,4]$时

$$t=\frac{1}{3}(2z-5),$$

$$y_2(z)=2.700\ 5+0.023\ 2\times\frac{1}{3}(2z-5)+\frac{2.316\ 0}{9}(2z-5)^2$$

$$=9.095\ 2-5.131\ 2z+1.029\ 3z^2$$

(7)计算总相对误差

$$d=\frac{\sum_{i=1}^{7}|y_i-y_2(z_i)|}{\sum_{i=1}^{7}|y_2(z_i)|}=\frac{0.486\ 9}{26.107\ 3}=0.018\ 65$$

由上述可见,通过线性变换,把$[a,b]$映射到$[-1,1]$上,然后利用切比雪夫多项式作为过渡,来确定最小二乘解的待定系数,从而避免了解一个病态矩阵的线性方程组。实质上,由于切比雪夫多项式的正交性,避免了解一个线性方程组,而直接得到最佳平方逼近的待定系数$a_k(k=0,1,2,\cdots,n)$。

依照定理要求,若想用各次切比雪夫多项式组成的线性无关组$\{T_k(x)\}(k=0,1,2,\cdots,n)$的线性组合构造$n$次最小二乘多项式,则必须求出$m$次$(m>n)$切比雪夫多项式的$m$个零点

$$x_j=\cos\frac{(2j-1)}{2m}\pi$$

这个次数m(注意m不是实验点数),只要在$m>n$的条件下,可大可小,究竟为多

小，自已确定。它只影响在确定多项式的系数 a_k 时求和的项数。若要想尽量减少求和时的计算次数，一般就取最小的 m，即 $m=n+1$。如例6－5中，寻求二次最小二乘法多项式，取 $m=3$。

第七节　多元线性最小二乘法

在前边各节中，较详细地讨论了一元函数的最小二乘法，在那里所建立的概念与给出的方法可直接推广到多元问题上去。

对于多变量离散函数 $(x_{ij},y_j)(i=1,2,\cdots,p;j=1,2,\cdots,m)$，利用线性最小二乘法拟合为多元函数

$$y=\sum_{k=1}^{n}a_kf_k(x_1,x_2,\cdots,x_p)$$

在式中，待定系数 $a_k(k=0,1,2,\cdots,n)$ 与 y 呈线性关系，对这类函数的最小二乘法称为线性最小二乘法。应指出的是在式中的 $f_k(x_1,x_2,\cdots,x_p)$ 不一定是线性的，并且往往是非线性的。上式只是其一般形式，下列几种函数可视为它的特殊形式。

(1)多元线性函数：$y=a_0x_0+a_1x_1+a_2x_2+\cdots+a_nx_n$；

(2)一元非线性函数：$y=a_0f_0(x)+a_1f_1(x)+a_2f_2(x)+\cdots+a_nf_n(x)$；

(3)一元多项式函数：$y=a_0+a_1x+a_2x^2+\cdots+a_nx^n$。

总之，只要待定系数 $a_k(k=0,1,2,\cdots,n)$ 与 y 呈线性关系的一切线性或非线性，一元或多元函数均可归入这一类，应用线性最少二乘法进行拟合。

化工技术中的大量半经验公式具有线性多元函数或非线性多元函数的形式，而且其中多数非线性多元函数可线性化为线性多元函数。例如四参数的饱和蒸气压方程 $\ln P=A+B/T+CT+D\ln T$，其中 A、B、C、D 为待定系数，便是一个可线性化的一元非线性函数。

现在讨论线性最小二乘法的一般解法。

设有离散函数 $(x_{ij},y_j)(i=1,2,\cdots,p;j=1,2,\cdots,m)$，现在线性无关函数组 $\{f_k(x_1,x_2,\cdots,x_p)\}$ 中找出某个线性组合

$$\hat{y}=a_1f_1(x_1,x_2,\cdots,x_p)+a_2f_2(x_1,x_2,\cdots,x_p)+\cdots+a_nf_n(x_1,x_2,\cdots,x_p)$$

来逼近它。这实际上就是求取待定系数 $a_k(k=1,2,\cdots,n)$ 使得能达到最佳平方逼近。

若自变量 $x=\{x_{ij}\}(i=1,2,\cdots,p;j=1,2,\cdots,m)$ 为已知的测定值，则函数的测定值 y_j 与估计值 $\hat{y}_j$ 之间依照最小二乘法原则应有

$$E=\min\sum_{j=1}^{m}\delta_j^2=\min\left[\sum_{j=1}^{m}\left(y_j-\sum_{k=1}^{n}a_kf_{kj}\right)^2\right]$$

于是应有

$$\frac{\partial E}{\partial a_k}=\frac{\partial\left[\sum_{j=1}^{m}\left(y_j-\sum_{k=1}^{n}a_kf_{kj}\right)^2\right]}{\partial a_k}=0\quad(k=1,2,\cdots,n)$$

即得求解 a_k 的线性方程组

$$a_1\sum_{j=1}^{m}f_{1j}f_{kj}+a_2\sum_{j=1}^{m}f_{2j}f_{kj}+\cdots+a_n\sum_{j=1}^{m}f_{nj}f_{kj}=\sum_{j=1}^{m}y_jf_{kj}\quad(k=1,2,\cdots,n)$$

写为矩阵形式即为

$$\begin{bmatrix}\sum_{j=1}^{m}f_{1j}f_{1j} & \sum_{j=1}^{m}f_{1j}f_{2j} & \cdots & \sum_{j=1}^{m}f_{1j}f_{nj}\\ \sum_{j=1}^{m}f_{2j}f_{2j} & \sum_{j=1}^{m}f_{2j}f_{2j} & \cdots & \sum_{j=1}^{m}f_{2j}f_{nj}\\ \vdots & \vdots & & \vdots\\ \sum_{j=1}^{m}f_{nj}f_{1j} & \sum_{j=1}^{m}f_{nj}f_{2j} & \cdots & \sum_{j=1}^{m}f_{nj}f_{nj}\end{bmatrix}\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix}=\begin{bmatrix}\sum_{j=1}^{m}y_jf_{1j}\\\sum_{j=1}^{m}y_jf_{2j}\\\vdots\\\sum_{j=1}^{m}y_jf_{nj}\end{bmatrix}$$

若记 $Y=(y_1,y_2,\cdots,y_n)$，用内积来表示和式为$(f_k,f_i)=\sum_{j=1}^{m}y_{kj}f_{ij}$，则上述方程组可写为

$$\begin{bmatrix}(f_1,f_1) & (f_1,f_2) & \cdots & (f_1,f_n)\\(f_2,f_1) & (f_2,f_2) & \cdots & (f_2,f_n)\\\vdots & \vdots & & \vdots\\(f_n,f_1) & (f_n,f_1) & \cdots & (f_n,f_n)\end{bmatrix}\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix}=\begin{bmatrix}(Y,f_1)\\(Y,f_2)\\\vdots\\(Y,f_n)\end{bmatrix}$$

由内积的交换性可知，方程组的系数矩阵 A 是对称矩阵。则可利用 LDL^{T} 分解法求解。解得$(a_1,a_2,\cdots,a_n)^{\mathrm{T}}$ 代入线性组合 $\hat{y}$即得所求的拟合函数

$$\hat{y}=a_1f_1(x_1,x_2,\cdots,x_p)+a_2f_2(x_1,x_2,\cdots,x_p)+\cdots+a_nf_n(x_1,x_2,\cdots,x_p)$$

应当注意的是离散函数的已知测定点的个数 m 应大于待定系数的个数 n，且对拟合的结果要进行显著性检验。

现记 $B=(a_1,a_2,\cdots,a_n)^{\mathrm{T}}$，则线性最小二乘法的正则方程组可定为 $AB=A_{n+1}$。于是其计算步骤可概括为：

(1)计算系数矩阵 A 的下三角元素及列向量。对于 $i=1,2,\cdots,n$，计算

$$a_{in+1}=\sum_{k=1}^{m}y_kf_{ik},$$

$$a_{ji}=\sum_{k=1}^{m}f_{jk}f_{ik}\quad(j=i,i+1,\cdots,n);$$

(2)利用 LDL^{T} 的分解法求 $B=(a_1,a_2,\cdots,a_n)^{\mathrm{T}}$；

(3)输出计算结果。

例 6-6 实验测得不同压力下纯水的沸点，试用线性最小二乘法求取四参数蒸气压方程

$$\ln p=a+bT+c/T+d\ln T$$

式中，P 为饱和蒸气压；T 为温度。实测数据为：

T/K	373.15	393.25	425.55	453.65	486.25	507.75	524.25	537.85	549.65
p/atm	1	2	5	10	20	30	40	50	60

解 该式为非线性一元函数，将其线性化后转化为多元线性回归问题，即令

$$f_1 = 1, \quad f_2 = T, \quad f_3 = \frac{1}{T}, \quad f_4 = \ln T, \quad y = \ln p$$

则

$$y = af_1 + bf_2 + cf_3 + df_4$$

于是，利用多元线性最小二乘法可得

$$a = 242.901\,9, \quad b = 0.037\,43, \quad c = -13\,382.471\,5, \quad d = -36.199\,0$$

最大相对误差

$$\max\left|\frac{\hat{p}_j - p_j}{\hat{p}_j}\right| = 1.37 \times 10^{-2}$$

总相对误差

$$d = \left|\frac{\sum_{i=1}^{9}|p_j - \hat{p}_j|}{\sum_{i=1}^{9}|\hat{p}_j|}\right| = \frac{875.61}{166\,186.13} = 5.27 \times 10^{-3}$$

可见拟合效果良好。故在 100～276 ℃之间水的饱和蒸气压计算式为

$$\ln p = 242.901\,9 + 0.037\,43T - \frac{13\,382.471\,5}{T} - 36.199\,0\ln T$$

第八节 显著性检验

显著性检验就是拟合所得的逼近函数对被逼近的离散函数（列表函数）的表示能力，也即拟合效果。这可通过对逼近函数的统计检验来判断，对于线性最小二乘法，常用的检验方法有方差分析和残差分析。

1. 方差分析

方差分析是从整体上判断逼近函数对离散函数的适应性。对于离散函数 (x_{ij}, y_j) $(i = 1, 2, \cdots, p; j = 1, 2, \cdots, m)$，通常当自变量 $X = (x_1, x_2, \cdots, x_p)$ 发生变化时，y 也相应地发生变化，称其为响应 y 的波动。波动的原因有两方面：一是 y 与 X 之间存在的回归关系而引起的；另一个是由于实验误差或其他原因引起的。即使在相同的 $X = (x_1, x_2, \cdots, x_p)$ 下测得的 y 值也不尽相同，一般用实测值 y_j 与平均值 $\bar{y}$ 差的平方和来描述总的波动情况，即

$$S_{总} = \sum_{j=1}^{m} (y_j - \bar{y})^2$$

其中，$\bar{y} = \frac{1}{m}\sum_{j=1}^{m} y_j$，$S_{总}$ 称为总平方和。

$$S_{总} = \sum_{j=1}^{m} (y_j - \bar{y})^2 = \sum_{j=1}^{m} [(y_j - \hat{y}_j) + (\hat{y}_j - \bar{y})]^2$$

$$= \sum_{j=1}^{m} (y_j - \hat{y})^2 + \sum_{j=1}^{m} (\hat{y}_j - \bar{y})^2 + 2\sum_{j=1}^{m} (y_j - \hat{y})(\hat{y}_j - \bar{y})$$

可以证明 $\sum_{j=1}^{m} (y_j - \hat{y})(\hat{y}_j - \bar{y}) = 0$，则

$$S_{总} = \sum_{j=1}^{m} (y_j - \hat{y})^2 + \sum_{j=1}^{m} (\hat{y}_j - \bar{y})^2 = Q + U$$

其中，$\hat{y}_i = \sum_{k=1}^{n} a_k f_k(x_{1j}, x_{2j}, \cdots, x_{pj})(j = 1,2,\cdots,m)$；$Q = \sum_{j=1}^{m} (y_j - \hat{y})^2$ 称为残差平方和；$U = \sum_{j=1}^{m} (\hat{y}_j - \bar{y})^2$ 称为回归平方和。

这样 $S_{总}$ 由残差平方和 Q 与回归平方和 U 组成。U 反映了 y 与自变量 $X = (x_1, x_2, \cdots, x_p)$ 之间存在的回归关系所引起的波动；Q 也称为剩余平方和，它表现了回归关系之外的其他因素（如实验误差、逼近函数本身的缺欠等）对总波动的影响，它包含有缺欠平方和（逼近函数对离散函数的表现能力）和误差平方和（实验测定的精确性）。

为了确定误差平方和，一般是采用重复实验，在第 j 个条件下重复 m_j 次实验，则第 j 个条件下的误差平方和为

$$\sum_{j=1}^{m_j} (y_{ji} - \bar{y}_j)^2 = \sum_{j=1}^{m_j} y_{ji}^2 - m_j \bar{y}_j^2 \quad (j = 1,2,\cdots,m)$$

其中，$\bar{y}_j = \frac{1}{m_j}\sum_{i=1}^{m_j} y_{ji}^2$。

若重复进行实验的条件共有 P 个，各个条件下的误差平方和相加，即为总误差平方和

$$S_{误} = \sum_{j=1}^{p} \left(\sum_{i=1}^{m_j} y_{ji}^2 - m_j \bar{y}_j^2 \right)$$

这样，残差平方和减去总误差平方和便为缺欠平方和：

$$S_{缺} = Q - S_{误} = \sum_{j=1}^{m} (\bar{y}_j - \hat{y}_j)^2$$

即

$$S_{缺} = \sum_{j=1}^{m} (y_j - \hat{y}_j)^2 - \sum_{j=1}^{p} \left(\sum_{i=1}^{m_j} y_{ji}^2 - m_j \bar{y}_j^2 \right)$$

由于各种平方和的数值与参与求和的项数有关，项数越多提供的信息越多，应将

项数的影响排除。为此引入自由度的概念。因共有 m 个离散数据，在计算 $\bar{y}$ 时有一个约束条件，则总平方和 $S_{总}=\sum_{j=1}^{m}(y_j-\bar{y})^2$ 的自由度 $f_{总}=m-1$；若共有 n 个特定参数，回归平方和 $U=\sum_{j=1}^{m}(\hat{y}_j-\bar{y})^2$ 的自由度 $f_{回}=n-1$，则剩余平方和 $Q=\sum_{j=1}^{m}(y_j-\hat{y})^2=S_{总}-U$ 的自由度 $f_{残}=m-n$；在第 j 个条件下重复实验次数为 m_j，而在计算 $\bar{y}_j$ 时用去一个自由度，则第 j 个实验条件的误差平方和 $\sum_{i=1}^{m_j}y_{ji}^2-m_j\bar{y}_j^2$ 的自由度为 m_j-1。因此，总误差平方和 $S_{误}=\sum_{j=1}^{p}\left(\sum_{i=1}^{m_j}(y_{ji}^2-m_j\bar{y}_j^2)\right)$ 的自由度 $f_{误}=\sum_{j=1}^{p}(m_j-1)$，于是缺欠平方和 $S_{缺}=Q-S_{误}$ 的自由度为 $f_{缺}=m-n-\sum_{j=1}^{p}(m_j-1)$。现将各平方和与其自由度之间的关系列入表 6-2 中。（各平方和除以相应的自由度称为均方和）

表 6-2

平方和	计算式	自由度	均方和	F	F_α
总	$S_{总}=\sum_{j=1}^{m}(y_j-\bar{y})^2$	$f_{总}=m-1$			
回归	$U=\sum_{j=1}^{m}(\hat{y}_j-\bar{y})^2$	$f_{回}=n-1$			
残差	$Q=S_{总}-U$	$f_{残}=m-n$			
误差	$S_{误}=\sum_{j=1}^{p}\left(\sum_{i=1}^{m_j}(y_{ji}^2-m_j\bar{y}_j^2)\right)$	$f_{误}=\sum_{j=1}^{p}(m_j-1)$	$\sigma_{误}=\frac{S_{误}}{f_{误}}$	$F=\frac{\sigma_{缺}}{\sigma_{误}}$	$F_\alpha(f_{缺},f_{误})$
缺欠	$S_{缺}=Q-S_{误}$	$f_{缺}=m-n-\sum_{j=1}^{p}(m_j-1)$	$\sigma_{缺}=\frac{S_{缺}}{f_{缺}}$		

应用缺欠均方和与误差均方和之比值 F 来判断逼近函数适应与否，进行 F 检验：

$$F=\frac{\text{缺欠均方和}}{\text{误差均方和}}=\frac{S_{缺}/f_{缺}}{S_{误}/f_{误}}$$

即

$$F=\frac{\left[\sum_{j=1}^{m}(\bar{y}_j-\hat{y}_j)^2\right]/\left[m-n-\sum_{j=1}^{p}(m_j-1)\right]}{\left[\sum_{j=1}^{p}\left(\sum_{i=1}^{m_j}y_{ji}^2-m_j\bar{y}_j^2\right)\right]/\left[\sum_{j=1}^{p}(m_j-1)\right]}$$

若

$$F<F_\alpha\left(m-n-\sum_{j=1}^{p}(m_j-1),\sum_{j=1}^{p}(m_j-1)\right)$$

则称逼近函数是显著的，表明逼近函数对离散函数在整体上有较好的表现能力，拟合得到的公式可供使用。其中 $F_\alpha(f_{缺},f_{误})$ 可由 F 分布表查得。在化工计算中，可取信度 α 为 0.05，0.025 或 0.01。

2. 残差分析

应用残差分析可找出逼近函数的局部缺陷，并加以纠正。即作出残差

$$\delta_i = y_j - \hat{y}_i \quad (j=1,2,\cdots,m)$$

与各自变量 $x_{ij}(i=1,2,\cdots,p;j=1,2,\cdots,m)$ 的图形，观察图形及 δ 的分布，如图 6-4 所示。

对每个变量均可画出残差分布图。图 6-4 中的(a)为残差的正态分布，说明逼近函数无缺陷，是正确的；(b)为残差分布随 x_i 的增大而发散，则应对逼近函数进行修正，对自变量取对数；(c)为残差随自变量 x_i 呈直线分布，则应在逼近函数中增加 x_i 的一个一次项；(d)为残差随 x_i 呈抛物形分布，则应在逼近函数中增加 x_i 的一个二次项。

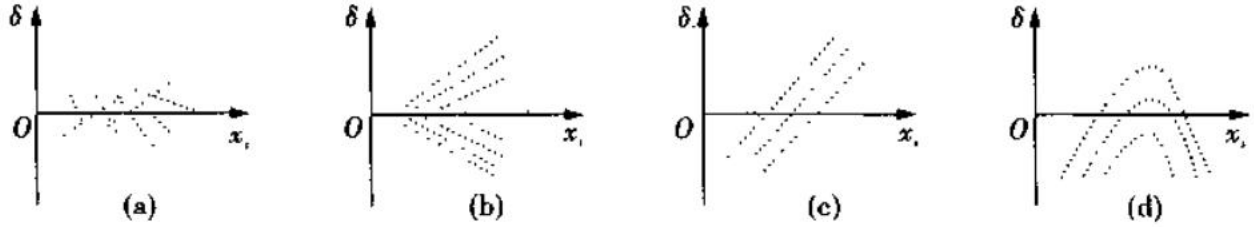


图 6-4

对逼近函数修正后再利用最小二乘法重新进行拟合，便可逐步求得适宜的逼近函数。

习 题

1. 证明切比雪夫多项式的以下恒等式：

(1) $T_m[T_n(x)] = T_{mn}(x)$；

(2) $T_{m+n}(x) + T_{m-n}(x) = 2T_m(x)T_n(x)$。

2. 设 $T_n^*(x) = T_n(2x-1)$，$x \in [0,1]$，求 $T_2^*(x)$。

3. 证明由第 2 题中的多项式组成的函数 $\{T_n^*(x)\}$ 是 $[0,1]$ 上带权 $\rho(x) = \dfrac{1}{\sqrt{x-x^2}}$ 的正交函数族。

4. 已知数据如下，试用最小二乘法求拟合函数：

x_i	0	1	2	3
y_i	0	0.500	0.866	1.00

5. 给定一组数据，用 $y = a\exp(bx)$ 函数进行拟合，求出待定常数 a、b。

x_i	1	2	3	4	5	6	7	8
y_i	15.3	20.5	27.4	36.6	49.1	65.6	87.8	117.6

6. 已知下述实验数据，试用切比雪夫多项式求其二次多项式的拟合曲线。

x_i	19	25	31	38	44
y_i	19.0	32.3	49.0	73.3	97.8

7. 已知异丁烷的饱和蒸气压和温度数据，试用线性最小二乘法确定方程 $\ln p = a + bT + c/T + d\ln T$ 中的参数 a、b、c、d。

T/K	188.06	201.45	216.72	229.04	245.57	251.08	254.39	259.91	261.54
p/mmHg	11.37	31.96	86.85	174.26	391.02	498.08	572.67	716.13	763.44

第七章　数值微分与数值积分

所谓函数$f(x)$的数值微分与数值积分是指对$f(x)$的近似函数$p(x)$进行微分与积分，从而得到$f(x)$的各阶导数与积分的近似值。所用到的近似函数主要指的是$f(x)$的各种插值多项式$p_n(x)$。特别是对于用列表方式给出的函数，只能用数值的方法求其微分与积分。

第一节　数值微分

数值微分就是求数值导数。例如化学反应动力学实验中，可直接测定出不同时间的反应物或产物的浓度$c(t)$，然后计算出在某一时刻浓度的变化率$\frac{dc(t)}{dt}$。这类问题就是数值微分问题。

由于多项式有较好的微分性能，原则上讲，利用插值多项式来确定函数$f(x)$的数值导数是没有问题的。但应该注意的是所得数值导数的误差估计，这是因为两个数值非常接近的函数，它们的导数可能差异很大。现将实际中常用的数值微分方法分述如下。

一、用差商近似微商

在微积分中，微商的概念为

$$f'(x)=\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=\lim_{h\to 0}\frac{f(x)-f(x-h)}{h}$$

$$=\lim_{h\to 0}\frac{f(x+\frac{h}{2})-f(x-\frac{h}{2})}{h}$$

则取其达到极限以前的形式就得到微商的差商近似式

$$f'(x)\approx\frac{f(x+h)-f(x)}{h}\approx\frac{f(x)-f(x-h)}{h}$$

$$\approx\frac{f(x+\frac{h}{2})-f(x-\frac{h}{2})}{h}$$

上式中3种不同表示形式依次是以一阶向前差商、一阶向后差商和一阶中心差商来近似表示微商。利用泰勒公式可看出这3种近似表达式的截断误差的数量级。

$$f(x+h)=f(x)+hf'(x)+\frac{1}{2}h^2f''(\xi)\quad(x<\xi<x+h)$$

则

$$f'(x)=\frac{f(x+h)-f(x)}{h}-\frac{1}{2}hf''(\xi)=\frac{\Delta f(x)}{h}+o(h)$$

$$f(x-h)=f(x)-hf'(x)+\frac{h^2}{2}f''(\eta)\quad(x-h<\eta<x)$$

则

$$f'(x)=\frac{f(x)-f(x-h)}{h}+\frac{h}{2}f''(\eta)=\frac{\nabla f(x)}{h}+o(h)$$

由中心差分定义可知

$$\begin{aligned}\delta f(x)&=f\left(x+\frac{h}{2}\right)-f\left(x-\frac{h}{2}\right)\\&=\left[f(x)+\frac{h}{2}f'(x)+\frac{1}{2}\left(\frac{h}{2}\right)^2f''(x)+\frac{1}{3!}\left(\frac{h}{2}\right)^3f'''(\xi_1)\right]-\\&\quad\left[f(x)-\frac{h}{2}f'(x)+\frac{1}{2}\left(\frac{h}{2}\right)^2f''(x)-\frac{1}{3!}\left(\frac{h}{2}\right)^3f'''(\xi_2)\right]\\&=hf'(x)+\frac{h^3}{48}[f'''(\xi_1)+f'''(\xi_2)]\end{aligned}$$

其中，$x<\xi_1<x+\frac{h}{2}$，$x-\frac{h}{2}<\xi_2<x$。

若记$f'''(\xi)=\frac{1}{2}[f'''(\xi_1)+f'''(\xi_2)]$，则

$$f'(x)=\frac{\delta f(x)}{h}-\frac{h^2}{24}f'''(\xi)=\frac{\delta f(x)}{h}+o(h^2)$$

由此可见，一阶向前差商和一阶向后差商近似表示微商的截断误差是关于步长h的一次方的高级无穷小量，而以一阶中心差商近似表示微商的截断误差是关于步长h的二次方的高级无穷小量，所以中心差商近似微商精度较高，从几何图形（如图7-1）上看，表现为弧段内接弦的斜率与切线斜率的平行程度在中点优于两端点。

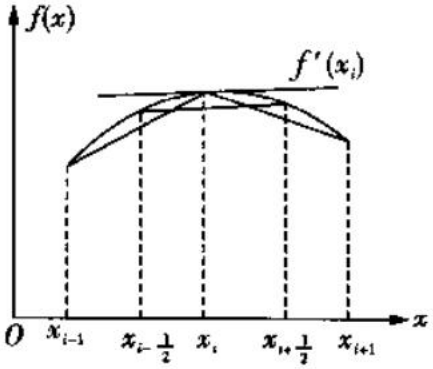


图7-1

对于二阶导数则有

$$f''(x)\approx\frac{\dfrac{\delta f\left(x+\frac{h}{2}\right)}{h}-\dfrac{\delta f\left(x-\frac{h}{2}\right)}{h}}{h}=\frac{f(x+h)-2f(x)+f(x-h)}{h^2}$$

$$=\frac{1}{h^2}\Delta^2 f(x-h)==\frac{1}{h^2}\nabla^2 f(x+h)$$

即

$$f''(x)=\frac{1}{h^2}\Delta^2 f(x-h)+o(h^2)$$

类似地可得

$$f'''(x)=\frac{1}{h^3}[f(x+2h)-3f(x+h)+3f(x)-f(x-h)]+o(h^2)$$

即

$$f'''(x)=\frac{1}{h^3}\Delta^3 f(x-h)+o(h^2)=\frac{1}{h^3}\nabla^3 f(x+2h)+o(h^2)$$

$$f^{(4)}(x)=\frac{1}{h^4}[f(x+2h)-4f(x+h)+6f(x)-4f(x-h)+f(x-2h)]+o(h^2)$$

即

$$f^{(4)}(x)=\frac{1}{h^4}\Delta^4 f(x-2h)+o(h^2)=\frac{1}{h^4}\nabla^4 f(x+2h)+o(h^2)$$

二、用插值函数计算微商

从插值法中可知，若给定函数 $y=f(x)$ 在插值区间 $[a,b]$ 上的 $n+1$ 个节点 $x_0,x_1,x_2,\cdots,x_n$，就意味着给出了一个列表函数 $(x_0,y_0),(x_1,y_1),\cdots,(x_n,y_n)$。这样就可构造一个 n 次插值多项式 $P_n(x)$ 来近似地代替 $f(x)$，即

$$f(x)=P_n(x)+R_n(x)\quad f(x)\approx P_n(x)$$

自然希望建立近似等式

$$f'(x)\approx P'_n(x)$$

类似地，也希望建立高阶微分的近似等式

$$f^{(k)}(x)\approx P_n^{(k)}(x)\quad (k=2,3,\cdots,n)$$

这类近似等式的误差究竟如何，这里将只对一阶导数的情况作出分析，并从中得出使用 $f'(x)\approx P'_n(x)$ 时应注意的问题。

由于

$$f(x)=P_n(x)+R_n(x)$$
$$f'(x)=P'_n(x)+R'_n(x)$$

又

$$R_n(x)=\frac{1}{(n+1)!}f^{(n+1)}(\xi)\cdot\omega(x)$$

其中，$\xi\in[a,b]$，并依赖于 x；$\omega(x)=\prod_{i=0}^{n}(x-x_i)$。

则

$$R'_n(x)=\frac{1}{(n+1)!}f^{(n+1)}(\xi)\cdot\omega'(x)+\frac{\omega(x)}{(n+1)!}\frac{\mathrm{d}}{\mathrm{d}x}[f^{(n+1)}(\xi)]$$

这是一个比较复杂的表达式。特别是第二项，涉及到$f(x)$的$n+2$阶导数的存在问题。即使假设$f^{(n+2)}(\xi)$存在，由于无法知道ξ本身依赖于x的具体形式，也无法求出它的值。然而，在一种重要的特殊情况下，它可得到简化。这就是若只限于求节点$x_i(i=0,1,2,\cdots,n)$处的数值导数，由于，当$x=x_i$时，$\omega(x_i)=0$，即上式中的第二项变为零。则

$$R'_n(x_i)=\frac{f^{(n+1)}(\xi)}{(n+1)!}\omega'(x_i)\leqslant\frac{M}{(n+1)!}\omega'(x_i)$$

其中，$\omega'(x_i)=\prod_{\substack{j=0\\j\neq i}}^{n}(x_i-x_j)$，$M=\max_{a\leqslant x\leqslant b}|f^{(n+1)}(x)|$。

这就是为什么通常数值微分总是在节点处进行的道理。因此，利用插值函数求数值导数时，应注意节点的选取，即把要求导数的点作为插值节点。

对于非等距节点时，由$f'(x)=P'_n(x)$求取导数。

对于等距节点，同样可由$f'(x)=P'_n(x)$求取导数。然而，由于等距的方便性，现就一阶导数给出不同节点数的各种近似微商公式。

1. 两点求导公式

在等距节点的情况下，$x_k=x_0+kh(k=0,1,\cdots,n)$，当$n=1$时，即有两个点$x_0,x_1$，可得线性插值公式

$$P_1(x)=\frac{x-x_1}{x_0-x_1}y_0+\frac{x-x_0}{x_1-x_0}y_1$$

$$P'(x)=\frac{1}{h}(y_1-y_0)\quad x\in[x_0,x_1]$$

而余项

$$R'_n(x)=\frac{f''(\xi)}{2!}\omega'(x_k)\quad(k=0,1)$$

则

$$R'_1(x_0)=\frac{1}{2!}f''(\xi)(x_0-x_1)=-\frac{1}{2}h\cdot f''(\xi)$$

$$R'_1(x_1)=\frac{1}{2!}f''(\xi)(x_1-x_0)=-\frac{1}{2}h\cdot f''(\xi)$$

于是，就得到节点x_0,x_1上的导数公式

$$y'_0=\frac{1}{h}(y_1-y_0)-\frac{1}{2}hy''(\xi)$$

$$y'_1=\frac{1}{h}(y_1-y_0)+\frac{1}{2}hy''(\xi)$$

采用类似的方法，可求得多点公式如下。

2. 三点公式，$n=2$

$$y'_0=\frac{1}{2h}(-3y_0+4y_1-y_2)+\frac{1}{3}h^2y^{(3)}$$

$$y'_1=\frac{1}{2h}(-y_0+0+y_2)-\frac{1}{6}h^2y^{(3)}$$

$$y'_2=\frac{1}{2h}(y_0-4y_1+3y_2)+\frac{1}{3}h^2y^{(3)}$$

3. 四点公式，$n=3$

$$y'_0=\frac{1}{6h}(-11y_0+18y_1-9y_2+2y_3)-\frac{1}{4}h^3y^{(4)}$$

$$y'_1=\frac{1}{6h}(-2y_0-3y_1+6y_2-y_3)+\frac{1}{12}h^3y^{(4)}$$

$$y'_2=\frac{1}{6h}(y_0-6y_1+3y_2+2y_3)-\frac{1}{12}h^3y^{(4)}$$

$$y'_3=\frac{1}{6h}(-2y_0+9y_1-18y_2+11y_3)+\frac{1}{4}h^3y^{(4)}$$

4. 五点公式，$n=4$

$$y'_0=\frac{1}{12h}(-25y_0+48y_1-36y_2+16y_3-3y_4)+\frac{1}{5}h^4y^{(5)}$$

$$y'_1=\frac{1}{12h}(-3y_0-10y_1+18y_2-6y_3+y_4)-\frac{1}{20}h^4y^{(5)}$$

$$y'_2=\frac{1}{12h}(y_0-8y_1+0+8y_3-y_4)+\frac{1}{30}h^4y^{(5)}$$

$$y'_3=\frac{1}{12h}(-y_0+6y_1-18y_2+10y_3+3y_4)-\frac{1}{20}h^4y^{(5)}$$

$$y'_4=\frac{1}{12h}(3y_0-16y_1+36y_2-48y_3+25y_4)-\frac{1}{5}h^4y^{(5)}$$

5. 六点公式，$n=5$

$$y'_0=\frac{1}{60h}(-137y_0+300y_1-300y_2+200y_3-75y_4+12y_5)-\frac{1}{6}h^5y^{(6)}$$

$$y'_1=\frac{1}{60h}(-12y_0-65y_1+120y_2-60y_3+20y_4-3y_5)+\frac{1}{30}h^5y^{(6)}$$

$$y'_2=\frac{1}{60h}(3y_0-30y_1-20y_2+60y_3-15y_4+2y_5)-\frac{1}{60}h^5y^{(6)}$$

$$y'_3=\frac{1}{60h}(-2y_0+15y_1-60y_2+20y_3+30y_4-3y_5)+\frac{1}{60}h^5y^{(6)}$$

$$y'_4=\frac{1}{60h}(3y_0-20y_1+60y_2-120y_3+65y_4+12y_5)-\frac{1}{30}h^5y^{(6)}$$

$$y'_5=\frac{1}{60h}(-12y_0+75y_1-200y_2+300y_3-300y_4+137y_5)+\frac{1}{6}h^5y^{(6)}$$

6. 七点公式，$n=6$

$$y'_0=\frac{1}{60h}(-147y_0+360y_1-450y_2+400y_3-225y_4+72y_5-10y_6)+\frac{1}{7}h^6y^{(7)}$$

$$y'_1=\frac{1}{60h}(-10y_0-77y_1+150y_2-100y_3+50y_4-15y_5+2y_6)-\frac{1}{42}h^6y^{(7)}$$

$$y'_2=\frac{1}{60h}(2y_0-24y_1-35y_2+80y_3-30y_4+8y_5-y_6)+\frac{1}{105}h^6y^{(7)}$$

$$y'_3=\frac{1}{60h}(-y_0+9y_1-45y_2+0+45y_4-9y_5+y_6)-\frac{1}{140}h^6y^{(7)}$$

$$y'_4=\frac{1}{60h}(y_0-8y_1+30y_2-80y_3+35y_4+24y_5-2y_6)+\frac{1}{105}h^6y^{(7)}$$

$$y'_5=\frac{1}{60h}(-2y_0+15y_1-50y_2+100y_3-150y_4+77y_5+10y_6)-\frac{1}{24}h^6y^{(7)}$$

$$y'_6=\frac{1}{60h}(10y_0-72y_1+225y_2-400y_3+450y_4-360y_5+147y_6)+\frac{1}{7}h^6y^{(7)}$$

从以上各组公式可看出：

(1)节点处函数值系数之和为零。

(2)当插值节点数为奇数时，中心点的导数公式不仅计算简单，而且误差较小。所以只要有可能，应尽量选用这些公式。这些公式也叫中心微商公式。

下面列出常用的几个中心微商公式，其中为了表明公式的对称性，把各点的序标号写成对称的形式。

(1)$n=2$ 时，三点中心微商公式

$$y'_0=\frac{1}{2h}(y_1-y_{-1})-\frac{1}{6}h^2y^{(3)}$$

(2)$n=4$ 时，五点中心微商公式

$$y'_0=\frac{2}{3h}(y_1-y_{-1})-\frac{1}{12h}(y_2-y_{-2})+\frac{1}{30}h^4y^{(5)}$$

(3)$n=6$ 时，七点中心微商公式

$$y'_0=\frac{3}{4h}(y_1-y_{-1})-\frac{3}{20h}(y_2-y_{-2})+\frac{1}{60h}(y_3-y_{-3})-\frac{1}{140}h^6y^{(7)}$$

(4)$n=8$ 时，九点中心微商公式

$$y'_0=\frac{4}{5h}(y_1-y_{-1})-\frac{1}{5h}(y_2-y_{-2})+\frac{4}{105h}(y_3-y_{-3})-\frac{1}{280h}(y_4-y_{-4})+\frac{h^8}{630}y^{(9)}$$

例 7-1 已知如下数据，试用三点、五点和七点中心微商公式求 $x=0.3$ 处的导数值。

x_i	0.0	0.1	0.2	0.3	0.4	0.5	0.6
y_i	1	0.995	0.980	0.955	0.921	0.878	0.825

解 由三点中心微商公式得

$$f'(0.3)=\frac{1}{2\times0.1}(0.921-0.980)=-0.295$$

由五点公式得

$$f'(0.3)=\frac{2}{3\times0.1}(0.921-0.980)-\frac{1}{12\times0.1}(0.878-0.995)$$
$$=-0.2958$$

由七点中心微商公式得

$$f'(0.3)=\frac{3}{4\times0.1}(0.921-0.980)-\frac{3}{20\times0.1}(0.878-0.995)+\frac{1}{60\times0.1}(0.825-1)$$
$$=-0.2962$$

例 7-1 中数值实际上是余弦函数在各点的近似值。该函数在 0.3 处的导数值为 -0.295 52。由计算的结果可知，插值节点增多并不一定能提高数值导数的精度，在实际计算中应注意这一点。如本例中就是用五点中心微商公式计算的结果的精度最高。

三、用三次样条函数求微商

应用三次样条插值函数 $s(x)$ 作为表格函数 $f(x)$ 的近似表达式，不但可使函数值非常接近，而且可使导数值也很接近。因为当 $f(x)$ 具有连续四阶导数，且 $h=\max_{1\leqslant i\leqslant n}h_i$ 趋于零时，$s(x)$，$s(x)'$，$s''(x)$，$s'''(x)$ 分别收敛于 $f(x)$，$f'(x)$，$f''(x)$，$f'''(x)$，并且有

$$|f(x)-s(x)|=o(h^4),\quad |f'(x)-s'(x)|=o(h^3),$$
$$|f''(x)-s''(x)|\approx o(h^2),\quad |f'''(x)-s'''(x)|=o(h)$$

因此，应用三次样条插值函数求数值导数是可靠的，不但可求节点处的导数，而且可求非节点处的导数。这是工程计算中求取数值的有效方法。

对于以节点处二阶导数表示的三次样条函数为

$$f(x)\approx s(x)=\frac{(x_{k+1}-x)^3}{6h_k}M_k+\frac{(x-x_k)^3}{6h_k}M_{k+1}+\left(y_k-\frac{h_k^2}{6}M_k\right)\frac{(x_{k+1}-x)^3}{h_k}+$$
$$\left(y_{k+1}-\frac{h_k^2}{6}M_{k+1}\right)\frac{(x-x_k)}{h_k}$$

其中，$k=0,1,2,\cdots,n-1$；$x\in[x_k,x_{k+1}]$。

于是

$$f'(x)=s'(x)=-\frac{(x_{k+1}-x)^2}{2h_k}M_k+\frac{(x-x_k)^2}{2h_k}M_{k+1}+\frac{y_{k+1}-y_k}{h_k}-\frac{1}{6}h_k(M_{k+1}-M_k)$$

$$f''(x)=s''(x)=\frac{1}{h_k}M_k(x_{k+1}-x)+\frac{1}{h_k}M_{k+1}(x-x_k)$$

因此，节点处的一阶和二阶导数分别为

$$f'(x_k)=s'(x_k)=-\frac{h_k}{2}M_k-\frac{h_k}{6}(M_{k+1}-M_k)+\frac{y_{k+1}-y_k}{h_k}$$

$$f''(x) = s''(x) = M_k$$

例 7-2 某液体冷却时，温度随时间的变化数据为：

t/min	0	1	2	3	4	5
T/℃	92.0	85.3	79.5	74.5	70.2	67.0

试分别计算 $t=2$、3、4min 及 $t=1.5$、2.5、4.5 min 时的降温速率。

解 由题意可知，前者是计算节点处的一阶导数，后者是计算非节点处的一阶导数。

首先由图解法求得 $\left.\frac{dT}{dt}\right|_{t=0} = -7.15$，$\left.\frac{dT}{dt}\right|_{t=5} = -2.52$。则得

$$\begin{bmatrix} 2 & 1 & & & & \\ 0.5 & 2 & 0.5 & & & \\ & 0.5 & 2 & 0.5 & & \\ & & 0.5 & 2 & 0.5 & \\ & & & 0.5 & 2 & 0.5 \\ & & & & 1 & 2 \end{bmatrix} \begin{bmatrix} M_0 \\ M_1 \\ M_2 \\ M_3 \\ M_4 \\ M_5 \end{bmatrix} = \begin{bmatrix} 2.7 \\ 2.7 \\ 2.4 \\ 2.1 \\ 3.3 \\ 4.08 \end{bmatrix}$$

由追赶法解得

$$M_0 = 0.8902, \quad M_1 = 0.9197, \quad M_2 = 0.831,$$
$$M_3 = 0.5564, \quad M_4 = 1.1438, \quad M_5 = 1.4681$$

对节点处

$$f'(x_k) = s'(x_k) = -\frac{1}{2}M_k - \frac{1}{6}(M_{k+1} - M_k) + y_{k+1} - y_k$$

$$= -\frac{1}{6}(2M_k + M_{k+1}) + y_{k+1} - y_k$$

对非节点处

$$f'(x) = s'(x) = -\frac{M_k}{2}(x_{k+1} - x)^2 + \frac{M_{k+1}}{2}(x - x_k)^2 + y_{k+1} - y_k - \frac{1}{6}(M_{k+1} + M_k)$$

代入相应数据计算得：

t/min	2	3	4	1.5	2.5	4.5
T/℃	79.5	74.5	70.2	82.29	76.91	68.44
$\frac{dT}{dt}$/(℃ · min^{-1})	-5.37	-4.68	-3.83	-5.80	-4.99	-3.21

第二节　数值积分

在积分学中，要计算函数$f(x)$在$[a,b]$上的定积分$I=\int_a^b f(x)\mathrm{d}x$，通常是先求得函数$f(x)$的一个原函数$F(x)$，再利用牛顿—莱布尼兹公式

$$I=\int_a^b f(x)\mathrm{d}x=F(b)-F(a)$$

来确定积分值。然而，这种方法仅适用于$f(x)$有解析表达式且原函数$F(x)$易于求得的情况。对于不能得到原函数，甚至连解析表达式都不存在的情况，例如表格函数等，就不能用其来确定积分值了，这类问题的解决有赖于数值积分。

在化工设备的工艺计算中，经常遇到求定积分的问题。例如，求取固定床催化剂用量；间歇反应器中达到一定转化率时所需要的反应时间；热力学中的逸度系数计算以及简单精馏塔高度计算等都是求定积分的问题，而且在这些积分中，被积函数往往比较复杂，甚至以离散形式给出，因而也就无法求其解析解，只能进行数值积分或图解积分。

数值积分的基本设想是通过一批给定的点及相应的函数值，来确定函数在$[a,b]$上的积分值的近似值。

设在$n+1$个插值节点上给出一个列表函数，即$(x_0,y_0),(x_1,y_1),\cdots,(x_n,y_n)$，可以构造一个$n$次插值多项$P_n(x)$来近似地代替$f(x)$，再对多项式$P_n(x)$进行积分，即得到函数$f(x)$在$[a,b]$上积分的近似值

$$I=\int_a^b P_n(x)\mathrm{d}x$$

显然，这种近似替代的形式并不是唯一的，按照不同的插值公式就可以得到不同的近似公式。这种求积分近似值的方法就叫数值积分法。

常用的插值型求积分公式有两类：一类是等距节点的牛顿—柯特斯求积公式；另一类是不等距节点的高斯型求积公式。下面就分别对其进行讨论。

第三节　牛顿—柯特斯求积公式

一、牛顿—柯特斯求积公式形式

牛顿—柯特斯求积公式的思想是用拉格朗日插值多项式$P_n(x)$作为被积函数$f(x)$的近似式来求取

$$I=\int_a^b f(x)\mathrm{d}x=\int_a^b P_n(x)\mathrm{d}x$$

将积分区间$[a,b]$$n$等分，即$h=(b-a)/n$，于是就有

$$x_k=a+kh\quad(k=0,1,2,\cdots,n)$$

$$I = \int_a^b f(x)\mathrm{d}x = \int_a^b \sum_{k=0}^{n} y_k L_k(x)\mathrm{d}x = \sum_{k=0}^{n} y_k \int_a^b \frac{\omega(x)}{(x-x_k)\omega'(x_k)}\mathrm{d}x$$

若记

$$A_k = \int_a^b \frac{\omega(x)}{(x-x_k)\omega'(x_k)}\mathrm{d}x \quad (k=0,1,\cdots,n)$$

则

$$I = \int_a^b f(x)\mathrm{d}x = \int_a^b P_n(x)\mathrm{d}x = \sum_{k=0}^{n} y_k A_k$$

作变量代换

$$x = a + ht, \quad \mathrm{d}x = h\mathrm{d}t \quad (\text{当 } x=a \text{ 时}, t=0; \text{当 } x=b \text{ 时}, t=n)$$

而

$$\omega(x) = \prod_{j=0}^{n}(x-x_j) = h^{n+1}t(t-1)(t-2)\cdots(t-n)$$

$$\omega'(x_k) = \prod_{\substack{j=0 \\ j\neq k}}^{n}(x_k-x_j) = (-1)^{(n-k)}h^n k!(n-k)! \quad (k=0,1,\cdots,n)$$

$$x - x_k = h(t-k)$$

则

$$\begin{aligned} A_k &= \int_0^n \frac{h^{n+1}t(t-1)(t-2)\cdots(t-n)h}{h(t-k)(-1)^{(n-k)}h^n k!(n-k)!}\mathrm{d}t \\ &= \frac{(-1)^{(n-k)}}{k!(n-k)!}h\int_0^n \prod_{\substack{j=0 \\ j\neq k}}^{n}(t-j)\mathrm{d}t \\ &= \frac{(-1)^{(n-k)}}{n\cdot k!(n-k)!}(b-a)\int_0^n \prod_{\substack{j=0 \\ j\neq k}}^{n}(t-j)\mathrm{d}t \\ &= (b-a)\cdot C_k^{(n)} \quad (k=0,1,\cdots,n) \end{aligned}$$

其中，$C_k^{(n)} = \dfrac{(-1)^{(n-k)}}{n\cdot k!\ (n-k)!}\displaystyle\int_0^n \prod_{\substack{j=0 \\ j\neq k}}^{n}(t-j)\mathrm{d}t \quad (k=0,1,\cdots,n)$。

$C_k^{(n)}$ 称为柯特斯系数，它只依赖于节点数 n，而与被积函数和积分区间均无关。因此，只要给定 $n=\dfrac{1}{h}(b-a)$，便可求出柯特斯系数 $C_k^{(n)}(k=0,1,\cdots,n)$。

于是

$$I = \int_a^b f(x)\mathrm{d}x = \int_a^b P_n(x)\mathrm{d}x = \sum_{k=0}^{n} y_k A_k = (b-a)\sum_{k=0}^{n} C_k^{(n)} \cdot f(a+kh)$$

该式称作牛顿—柯特斯求积公式，也称为等距节点求积公式。

这样，对同一积分，在$[a,b]$上取 n 不同，则节点

$$x_k = a + k\frac{b-a}{n} \quad (k=0,1,\cdots,n)$$

的数目及分布就不同，因而数值积分的精度也不同。

当 $n=1$ 时，$h=b-a, x_0=a, x_1=b$，于是

$$C_0^{(1)} = -\int_0^1 (t-1)\,dt = \frac{1}{2}$$

$$C_1^{(1)} = \int_0^1 t\,dt = \frac{1}{2}$$

于是

$$\int_a^b f(x)\,dx = \frac{b-a}{2}[f(a)+f(b)]$$

上式称为梯形公式，其几何意义就是由梯形而积近似地代替曲边梯形的面积，见图 7-2(a)。

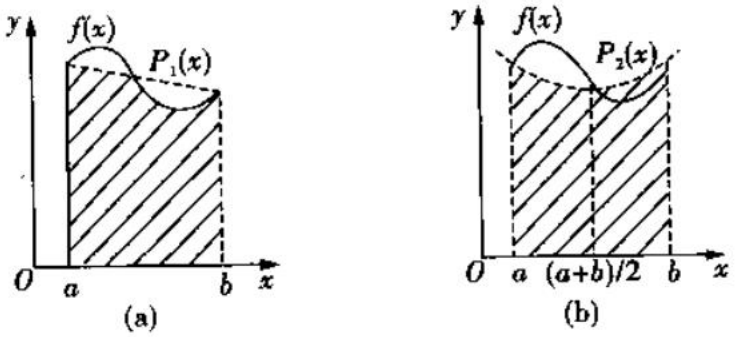


图 7-2

当 $n=2$ 时，$h=\frac{1}{2}(b-a), x_0=a, x_1=\frac{a+b}{2}, x_2=b$，有

$$C_0^{(2)} = \frac{1}{4}\int_0^2 (t-1)(t-2)\,dt = \frac{1}{6}$$

$$C_1^{(2)} = -\frac{1}{2}\int_0^2 t(t-2)\,dt = \frac{4}{6} = \frac{2}{3}$$

$$C_2^{(2)} = \frac{1}{4}\int_0^2 t(t-1)\,dt = \frac{1}{6}$$

则

$$\int_a^b f(x)\,dx \approx \frac{b-a}{6}\left[f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right]$$

该式称为辛普生公式，其几何意义就是用抛物线下的面积，近似代替曲线下的面积，见图 7-2(b)。

当 $n=3$ 时，$h=\frac{1}{3}(b-a), x_k=a+\frac{k}{3}(b-a)\quad (k=0,1,2,3)$，同样方法可求得

$$C_0^{(3)}=\frac{1}{8},\quad C_1^{(3)}=\frac{3}{8},\quad C_2^{(3)}=\frac{3}{8},\quad C_3^{(3)}=\frac{1}{8}$$

则

$$\int_a^b f(x)\,dx = \frac{b-a}{8}[f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)]$$

此式称为辛普生$\frac{3}{8}$法则。

当 $n=4$ 时，$x_k = a + \frac{k}{4}(b-a)$ $(k=0,1,2,3,4)$，可求得

$$C_0^{(4)} = \frac{7}{90},\quad C_1^{(4)} = \frac{16}{45},\quad C_2^{(4)} = \frac{2}{15},\quad C_3^{(4)} = \frac{16}{45},\quad C_4^{(4)} = \frac{7}{90}$$

则

$$\int_a^b f(x)\,dx \approx \frac{b-a}{90}[7f(x_0) + 32f(x_1) + 12f(x_2) + 32f(x_3) + 7f(x_4)]$$

此式称作柯特斯公式。

如此可求得一系列积分近似值。一般来说，随着 n 的增加，近似积分的精度提高。但当 $n \geq 8$ 后，柯特斯系数出现负值，使得计算不稳定，不能使用。为提高精度，可采用其他的算法。表 7－1 列出了不同 n 值下的柯特斯系数。

表 7－1　柯特斯系数

n	$C_k^{(n)}$								
1	$\frac{1}{2}$	$\frac{1}{2}$							
2	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$						
3	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$					
4	$\frac{7}{90}$	$\frac{16}{45}$	$\frac{2}{15}$	$\frac{16}{45}$	$\frac{7}{90}$				
5	$\frac{19}{288}$	$\frac{25}{96}$	$\frac{25}{144}$	$\frac{25}{144}$	$\frac{25}{96}$	$\frac{19}{288}$			
6	$\frac{41}{840}$	$\frac{9}{35}$	$\frac{9}{280}$	$\frac{34}{105}$	$\frac{9}{280}$	$\frac{9}{35}$	$\frac{41}{840}$		
7	$\frac{751}{17\,280}$	$\frac{3\,577}{17\,280}$	$\frac{1\,323}{17\,280}$	$\frac{2\,989}{17\,280}$	$\frac{2\,989}{17\,280}$	$\frac{1\,323}{17\,280}$	$\frac{3\,577}{17\,280}$	$\frac{751}{17\,280}$	
8	$\frac{989}{28\,350}$	$\frac{5\,888}{28\,350}$	$\frac{-928}{28\,350}$	$\frac{10\,496}{28\,350}$	$\frac{-4\,540}{28\,350}$	$\frac{10\,496}{28\,350}$	$\frac{-928}{28\,350}$	$\frac{5\,888}{28\,350}$	$\frac{989}{28\,350}$

二、牛顿—柯特斯公式的精度

表示插值型求积公式精度的方法有两种：一种为截断误差，即余项；另一种为代数精度。余项表示求积公式的近似程度，代数精度表明它适用的广度。

1. 截断误差

由于

$$R_n(x) = f(x) - P_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\omega(x)$$

则

$$R_n[f] = \int_a^b R_n(x)\mathrm{d}x = \frac{1}{(n+1)!}\int_a^b f^{(n+1)}(\xi)\cdot\omega(x)\mathrm{d}x$$

其中，$\xi \in [a,b]$ 且依赖于 x。

对于 $n=1$ 的梯形公式，则有

$$R_1[f] = \frac{1}{2}\int_a^b f''(\xi)(x-a)(x-b)\mathrm{d}x$$

由于 $(x-a)(x-b)$ 在区间 $[a,b]$ 上的符号保持不变，若 $f''(x)$ 连续，那么由积分中值定理可知

$$R_1[f] = \frac{1}{2}\int_a^b f''(\xi)(x-a)(x-b)\mathrm{d}x = \frac{1}{2}f''(\eta)\int_a^b (x-a)(x-b)\mathrm{d}x$$

$$= -\frac{1}{12}(b-a)^3 f''(\eta)$$

其中，$\eta \in [a,b]$ 且依赖于 x。

由此式可估计梯形公式的误差限

$$|R_1[f]| = \frac{1}{12}(b-a)^3 \max_{a\leqslant x\leqslant b}|f''(x)|.$$

对于其他的牛顿—柯特斯公式，可导出截断误差（余项）为

$$R_n[f] = \begin{cases} \dfrac{f^{(n+2)}(\eta)}{(n+2)!}\displaystyle\int_a^b x\omega(x)\mathrm{d}x & (n \text{ 为偶数时}) \\ \dfrac{f^{(n+1)}(\eta)}{(n+1)!}\displaystyle\int_a^b \omega(x)\mathrm{d}x & (n \text{ 为奇数时}) \end{cases}$$

其中，$\eta \in [a,b]$。

则对于辛普生公式，$n=2$，$\omega(x) = (x-a)\left(x-\dfrac{a+b}{2}\right)(x-b)$，

$$R_2[f] = \frac{f^{(4)}(\eta)}{4!}\int_a^b x(x-a)\left(x-\frac{a+b}{2}\right)(x-b)\mathrm{d}x = -\frac{(b-a)^5}{2\,880}f^{(4)}(\eta),$$

$$|R_2[f]| = \frac{(b-a)^5}{2\,880}\max_{a\leqslant x\leqslant b}|f^{(4)}(x)|;$$

当 $n=3$ 时，$\omega(x) = (x-a)\left(x-\dfrac{b+2a}{3}\right)\left(x-\dfrac{a+2b}{3}\right)(x-b)$，

$$R_3[f] = \frac{f^{(4)}(\eta)}{4!}\int_a^b (x-a)\left(x-\frac{b+2a}{3}\right)\left(x-\frac{a+2b}{3}\right)(x-b)\mathrm{d}x$$

$$= -\frac{(b-a)^5}{6\,480}f^{(4)}(\eta),$$

$$|R_3[f]| = \frac{(b-a)^5}{6\,480}\max_{a\leqslant x\leqslant b}|f^{(4)}(x)|;$$

同样,对于柯特斯公式,即 $n=4$,有

$$R_4[f] = -\frac{8}{945}\left(\frac{b-a}{4}\right)^7 f^{(6)}(\eta),$$

$$|R_4[f]| = \frac{(b-a)^7}{1\,935\,360}\max_{a\leqslant x\leqslant b}|f^{(6)}(x)|。$$

可见,只要 $f^{(n+1)}(x)$ 或 $f^{(n+2)}(x)$ 在区间 $[a,b]$ 上连续,即可计算 n 阶牛顿—柯特斯公式的余项,并估计其误差限。

2. 代数精度

定义 若求积公式 $\int_a^b f(x)\mathrm{d}x = \sum_{k=0}^{n} A_k f(x_k)$ 对于所有次数不高于 n 的多项式能准确成立,而对 $n+1$ 次多项式不能准确成立,则称该求积公式具有 n 次代数精度。

显然,求积公式的代数精度就是由线性组合式所能精确计算积分值的最高次多项式的次数。

现对梯形公式

$$\int_a^b f(x)\mathrm{d}x = \frac{1}{2}(b-a)[f(a)+f(b)]$$

当 $f(x)=x$,即 $f(x)$ 为一次多项式时,积分

$$\int_a^b f(x)\mathrm{d}x = \frac{1}{2}(b^2-a^2) = \frac{1}{2}(b-a)(b+a)$$

得梯形求积公式

$$\int_a^b x\mathrm{d}x = \frac{1}{2}(b-a)(b+a)$$

当 $f(x)=x^2$,即 $f(x)$ 为二次多项式时,积分

$$\int_a^b x^2\mathrm{d}x = \frac{1}{3}(b^3-a^3)$$

此时,梯形求积公式为

$$\int_a^b x^2\mathrm{d}x = \frac{1}{2}(b-a)(b^2+a^2) \neq \frac{1}{3}(b^3-a^3)$$

这说明梯形求积公式具有 1 次代数精度。

再看辛普生公式

$$\int_a^b f(x)\mathrm{d}x = \frac{b-a}{6}\left[f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right]$$

当 $f(x)=x$ 时,积分

$$\int_a^b x\mathrm{d}x = \frac{1}{2}(b^2-a^2)$$

此时,辛普生公式为

$$\int_a^b x\mathrm{d}x = \frac{b-a}{6}\left[a + 4\left(\frac{a+b}{2}\right) + b\right] = \frac{1}{2}(b^2 - a^2)$$

当$f(x) = x^2$时，积分

$$\int_a^b x^2\mathrm{d}x = \frac{1}{3}(b^3 - a^3)$$

此时，辛普生公式为

$$\int_a^b x^2\mathrm{d}x = \frac{b-a}{6}\left[a^2 + 4\left(\frac{a+b}{2}\right)^2 + b^2\right] = \frac{1}{3}(b^3 - a^3)$$

当$f(x) = x^3$时，积分

$$\int_a^b x^3\mathrm{d}x = \frac{1}{4}(b^4 - a^4)$$

此时，辛普生公式为

$$\int_a^b x^3\mathrm{d}x = \frac{b-a}{6}\left[a^3 + 4\left(\frac{a+b}{2}\right)^3 + b^3\right] = \frac{1}{4}(b^4 - a^4)$$

当$f(x) = x^4$时，积分

$$\int_a^b x^4\mathrm{d}x = \frac{1}{5}(b^5 - a^5)$$

此时，辛普生公式为

$$\int_a^b x^4\mathrm{d}x = \frac{b-a}{6}\left[a^4 + 4\left(\frac{a+b}{2}\right)^4 + b^4\right] \neq \frac{1}{5}(b^5 - a^5)$$

这说明辛普生公式具有3次代数精度。

同理可推得，对于n阶的牛顿—柯特斯公式，当n为奇数时，具有n次代数精度；当n为偶数时，至少具有$n+1$次代数精度。

代数精度越高的求积公式，就能使更多的被积函数$f(x)$的数值积分具有更小的截断误差，使其适用范围更宽。如辛普生公式$n=2$，柯特斯公式$n=4$，在数值积分中得到广泛的应用。

第四节　复化求积公式

牛顿—柯特斯公式是等距节点的插值型求积公式。一般来说，节点数n增多，可提高其阶数，使其代数精度提高，截断误差减小。但当n过大时，计算公式过于复杂，并且当$n \geq 8$时，柯特斯系数有正有负，使计算不稳定，收敛无保证，计算误差反而增大。因此，在实际计算中，并不采用n较大的求积公式，而是将积分区间$[a,b]$预先分成N个子区间，每个子区间的长度为$h=(b-a)/N$。然后对每个子区间施以n较小的求积公式求积，最后将每个子区间的近似积分值相加，便求得积分区间$[a,b]$上的总积分近似值。这种算法称作复化求积。下面给出几个常用的复化求积公式。

一、复化梯形公式

将积分区间$[a,b]$$N$等分，则每个子区间宽度为$h_T=(b-a)/N$，分点为$x_k=a+kh_T$$(k=0,1,\cdots,N)$。对每个子区间施以梯形公式，则得复化梯形公式

$$\begin{aligned}\int_a^b f(x)\,\mathrm{d}x &= \sum_{k=1}^{N}\int_{x_{k-1}}^{x_k} f(x)\,\mathrm{d}x \approx \frac{1}{2}h_T\sum_{k=1}^{N}[f(x_{k-1})+f(x_k)]\\ &= \frac{1}{2}h_T[f(x_0)+2\sum_{k=1}^{N-1}f(x_k)+f(x_N)]\\ &= \frac{b-a}{2N}\left[f(a)-f(b)+2\sum_{k=1}^{N}f\left(a+k\frac{b-a}{N}\right)\right]\\ &\equiv T_N\end{aligned}$$

该式的几何意义，就是利用N条直线所构成的折线代替曲线$y=f(x)$来求定积分。

复化梯形公式的截断误差可表示为

$$\begin{aligned}R_T[f] &= \int_a^b f(x)\,\mathrm{d}x - T_N = -\frac{1}{12}h_T^3[f''(\eta_1)+f''(\eta_2)+\cdots+f''(\eta_n)]\\ &= -\frac{1}{12}h_T^3Nf''(\eta) = -\frac{b-a}{12}h_T^2\cdot f''(\eta)\end{aligned}$$

其中，$f''(\eta)=\frac{1}{N}[f''(\eta_1)+f''(\eta_2)+\cdots+f''(\eta_n)]$；$\eta\in[a,b]$；$\eta_k\in[x_{k-1},x_k]$，$k=0,1,\cdots,N$。

当$N\to\infty$时，$h_T\to0$，T_N收敛于积分$\int_a^b f(x)\,\mathrm{d}x$，且收敛速度为$o(h_T^2)$，称为二阶收敛。只要$N$足够大，$T_N$的精度便可足够高。但是，当$N$太大时，函数$f(x)$$(k=0,1,2,\cdots,N)$的计算次数太多，累积误差的影响就增大，使$T_N$的计算精度受到限制。

二、复化辛普生公式

若在每个子区间上施以辛普生公式便可得复化辛普生公式。由于辛普生公式在每个子区间上都需要3个节点，即$x_{k-1},x_{k-\frac{1}{2}},x_k$（这里$x_{k-\frac{1}{2}}=\frac{x_{k-1}+x_k}{2}$）。因此，在$N$个子区间上共有节点$2N+1$个，即$x_k=a+\frac{1}{2}kh_T=a+kh_s(k=0,1,2,\cdots,N)$，其中包括分点$N+1$个，即$x_{2k}=a+kh_T=a+2kh_s(k=0,1,2,\cdots,N)$，式中，$h_s=h_T/2=\frac{1}{2N}(b-a)$称为节点间距。于是复化辛普生公式为

$$\begin{aligned}\int_a^b f(x)\,\mathrm{d}x &= \sum_{k=1}^{N}\int_{x_{2k-2}}^{x_{2k}} f(x)\,\mathrm{d}x \approx \frac{1}{6}h_T\sum_{k=1}^{N}[f(x_{2k-2})+4f(x_{2k-1})+f(x_{2k})]\\ &= \frac{1}{6}h_T[f(x_0)+4\sum_{k=1}^{N}f(x_{2k-1})+2\sum_{k=1}^{N-1}f(x_{2k})+f(x_{2N})]\end{aligned}$$

$$= \frac{1}{3}h_s\left[f(a) - f(b) + 4\sum_{k=1}^{N} f(a + (2k-1)h_s) + 2\sum_{k=1}^{N} f(a + 2kh_s)\right]$$

$$\equiv S_N$$

该式的几何意义就是用 N 条抛物线所构成的连线代替曲线 $y = f(x)$ 来求积分。

复化辛普生公式的截断误差为

$$R_s[f] = \int_a^b f(x)\,\mathrm{d}x - S_N = -\frac{h_T^5}{2\,880}\sum_{k=1}^{N} f^{(4)}(\eta_k) = -\frac{b-a}{2\,880}h_T^4 f^{(4)}(\eta)$$

其中，$\eta \in [a,b]$；$\eta_k \in [x_{k-1}, x_k]$，$k = 0,1,\cdots,N$；$f^{(4)}(\eta) = \frac{1}{N}\sum_{k=1}^{N} f^{(4)}(\eta_k)$。

可见，当 $N \to \infty$ 时，$h_T \to 0$，S_N 收敛于 $\int_a^b f(x)\,\mathrm{d}x$，且为四阶收敛，其收敛速度快于复化梯形公式。

三、复化柯特斯公式

若在每个子区间上施以柯特斯公式便得复化柯特斯公式。由于柯特斯公式在每个子区间上都需要 5 个节点，所以在 N 个子区间内共有 $4N+1$ 个节点，即

$$x_k = a + \frac{1}{4}kh_T = a + kh_c \quad (k = 0,1,2,\cdots,4N)$$

其中包括 $N+1$ 个分点，即

$$x_k = a + kh_T = a + 4kh_c \quad (k = 0,1,\cdots,N)$$

式中

$$h_c = \frac{1}{4}h_T = \frac{1}{4N}(b-a)$$

称为节点间距。于是，复化柯特斯公式为

$$\int_a^b f(x)\,\mathrm{d}x = \sum_{k=1}^{N}\int_{x_{4k-4}}^{x_{4k}} f(x)\,\mathrm{d}x$$

$$\approx \frac{h_T}{90}\left[7f(x_0) + 32\sum_{k=1}^{N} f(x_{4k-3}) + 12\sum_{k=1}^{N} f(x_{4k-2}) + 32\sum_{k=1}^{N} f(x_{4k-1}) + 14\sum_{k=1}^{N-1} f(x_{4k}) + 7f(x_{4N})\right]$$

$$= \frac{4}{90}h_c\left[7f(a) - 7f(b) + 32\sum_{k=1}^{N} f(a + (4k-3)h_c) + 12\sum_{k=1}^{N} f(a + (4k-2)h_c) + 32\sum_{k=1}^{N} f(a + (4k-1)h_c) + 14\sum_{k=1}^{N} f(a + 4kh_c)\right]$$

$$\equiv C_N$$

复化柯特斯公式的截断误差为

$$R_c[f] = \int_a^b f(x)\mathrm{d}x - C_N = -\frac{2}{945}(b-a)\left(\frac{h_T}{4}\right)^6 f^{(6)}(\eta), \quad \eta \in [a,b]$$

显然，当 $N\to\infty$ 时，$h_T\to 0$，C_N 收敛于积分 $\int_a^b f(x)\mathrm{d}x$，且为六阶收敛，其收速度更快。

例 7-3 应用复化梯形公式、复化辛普生公式和复化柯特斯公式分别计算积分 $\int_0^1 \frac{\sin x}{x}\mathrm{d}x$ 的近似值，要求按照复化辛普生公式的计算误差不超过 0.5×10^{-6}。

解 首先根据精度要求，由 $|R_s[f]|$ 来估计应将区间 $[0,1]$ 分为几个子区间。则应有

$$|R_s[f]| \leqslant \frac{b-a}{2\,880}\left(\frac{b-a}{N}\right)^4 \max_{a\leqslant x\leqslant b}|f^{(4)}(x)| \leqslant 0.5\times10^{-6}$$

于是

$$N^4 \geqslant \frac{2\times10^6}{2\,880}(b-a)^5 \max_{a\leqslant x\leqslant b}|f^{(4)}(x)| = \frac{10^6}{1\,440}\max_{a\leqslant x\leqslant b}\left|\left(\frac{\sin x}{x}\right)^{(4)}\right|$$

因为

$$\because f(x) = \frac{1}{x}\sin x = \int_0^1 \cos(xt)\mathrm{d}t$$

则

$$f^{(4)}(x) = \left(\frac{\sin x}{x}\right)^{(4)} = \int_0^1 \frac{\mathrm{d}^4}{\mathrm{d}x^4}\cos(xt)\mathrm{d}t = \int_0^1 t^4\cos(xt+2\pi)\mathrm{d}t$$

而

$$\max_{a\leqslant x\leqslant b}\left|\left(\frac{\sin x}{x}\right)^{(4)}\right| = \max_{a\leqslant x\leqslant b}\left|\int_0^1 t^4\cos(xt+2\pi)\mathrm{d}t\right| \leqslant \int_0^1 t^4\mathrm{d}t = \frac{1}{5}$$

即

$$N^4 \geqslant \frac{10^6}{1\,440}\times\frac{1}{5} = 138.89$$

$$N \geqslant (138.89)^{\frac{1}{4}} = 3.433$$

故取 $N=4$，也即将区间分为 4 个子区间，共需 9 个节点。

然后应用复化求积公式进行计算，应注意的是，对于复化梯形公式，9 个节点相当于将区间 $[0,1]$ 8 等分，而对复化柯特斯公式则相当于将区间 $[0,1]$ 2 等分。其计算结果分别为

$$T_8 = 0.945\,690\,531, \quad S_4 = 0.946\,083\,335, \quad C_2 = 0.946\,083\,093$$

而积分真值

$$I = \int_0^1 \frac{\sin x}{x}\mathrm{d}x = 1 - \frac{1}{3}\times\frac{1}{3!} + \frac{1}{5}\times\frac{1}{5!} - \frac{1}{7}\times\frac{1}{7!} + \cdots = 0.946\,083\,1$$

可见，复化辛普生公式和复化柯特斯公式有相当高的精度。

复化辛普生公式与复化梯形公式相比，收敛阶高二阶，代数精度高 2 次，而在节点数相同时，函数计算次数相同，计算量相当，但计算精度要高得多。因此，复化辛普生公式

在实际计算中得到广泛应用。

另外，还应指出的是，在利用复化求积公式进行计算时，当 N 较小时，随着 N 增大，计算精度提高；当 N 过大时，函数计算次数增多，舍入误差的积累反而使计算精度降低。通常，对于一定的精度要求，存在一个适宜的 N 值，这个 N 值可由 $|R_n[f]|$ 式来估计。但是，因在估计中要找 $\max\limits_{a\leqslant x\leqslant b}|f^{(n+1)}(x)|$ 或 $\max\limits_{a\leqslant x\leqslant b}|f^{(n+2)}(x)|$，所以这通常是很困难的。

第五节　加速求积公式

复化求积公式对提高精度是行之有效的，但要使用它必须先给出合适的步长，即要有一个适宜的 N，而适宜的 N 值通常是难于事先估计的。实际计算时，是采用变步长的积分法，即逐次将积分区间 $[a,b]$ 分半，$h=(b-a)/N$，而 $N=2^i(i=0,1,2,\cdots)$，并逐次地应用复化求积公式进行计算。若先后两次积分结果的差值满足要求，则最后一次的积分结果便为积分近似值。

一、变步长求积公式

若将 $[a,b]$ 分为 N 等份，则对复化梯形公式有

$$\begin{aligned}I&=\int_a^b f(x)\,\mathrm{d}x\\&=\frac{1}{2}h_T[f(a)-f(b)+2\sum_{k=1}^{N}f(a+kh_T)]-\frac{b-a}{12}\left(\frac{b-a}{N}\right)^2f''(\eta_1)\\&=T_N-\frac{b-a}{12}\left(\frac{b-a}{N}\right)^2f''(\eta_1)\end{aligned}$$

其中，$\eta_1\in[a,b]$。现将 $[a,b]$ 分为 $2N$ 等份，则又有

$$\begin{aligned}I&=\int_a^b f(x)\,\mathrm{d}x\\&=\frac{1}{4}h_T\left[f(a)-f(b)+2\sum_{k=1}^{2N}f\left(a+\frac{1}{2}kh_T\right)\right]-\frac{b-a}{12}\left(\frac{b-a}{2N}\right)^2f''(\eta_2)\\&=T_{2N}-\frac{b-a}{12}\left(\frac{b-a}{2N}\right)^2f''(\eta_2)\end{aligned}$$

其中，$\eta_2\in[a,b]$。

于是

$$\frac{I-T_N}{I-T_{2N}}=4\frac{f''(\eta_1)}{f''(\eta_2)}$$

若 $f''(x)$ 在 $[a,b]$ 上变化不大，取 $f''(\eta_1)\approx f''(\eta_2)$，则

$$\frac{I-T_N}{I-T_{2N}}\approx 4$$

即有

$$I = \int_a^b f(x)\,dx = \frac{4}{4-1}T_{2N} - \frac{1}{4-1}T_N$$

因此

$$|R_T[f]| = |I - T_{2N}| = \frac{1}{3}|T_{2N} - T_N|$$

故可用前后两次积分结果的差值来判定复化梯形公式计算结果的误差。

类似地，可以推导出变步长的辛普生公式和变步长的柯特斯公式：

$$I = \int_a^b f(x)\,dx = \frac{4^2}{4^2-1}S_{2N} - \frac{1}{4^2-1}S_N$$

$$I = \int_a^b f(x)\,dx = \frac{4^3}{4^3-1}C_{2N} - \frac{1}{4^3-1}C_N$$

由上面的式子可以看出，复化求积公式在区间$[a,b]$逐次分半后，先后两次积分结果的线性组合可获得精度更高的计算结果。更有趣的是组合$\left(\frac{4}{4-1}T_{2N} - \frac{1}{4-1}T_N\right)$恰恰是区间$[a,b]$$N$等份时的复化辛普生公式的计算结果，而表达式$\left(\frac{4^2}{4^2-1}S_{2N} - \frac{1}{4^2-1}S_N\right)$又正好是区间$[a,b]$$N$等分时复化柯特斯公式的计算结果，即低精度求积公式先后两次计算结果的线性组合可求得高精度求积公式的计算值，从而使变步长求积过程加速。

现在证明复化梯形公式先后两次积分结果的线性组合便是复化辛普生公式的积分值，即

$$S_N = \frac{4}{4-1}T_{2N} - \frac{1}{4-1}T_N$$

当$N=1$时

$$T_1 = \frac{1}{2}(b-a)[f(a)+f(b)]$$

$$T_2 = \frac{1}{4}(b-a)\left[f(a) + 2f\left(\frac{a+b}{2}\right) + f(b)\right]$$

$$\begin{aligned} I = \int_a^b f(x)\,dx &= \frac{4}{3}T_2 - \frac{1}{3}T_1 \\ &= \frac{1}{3}(b-a)\left[f(a) + 2f\left(\frac{a+b}{2}\right) + f(b)\right] - \frac{1}{6}(b-a)[f(a)+f(b)] \\ &= \frac{1}{6}(b-a)\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right] \\ &= S_1 \end{aligned}$$

则

$$S_1 = \frac{4}{3}T_2 - \frac{1}{3}T_1 = \frac{4}{4-1}T_2 - \frac{4}{4-1}T_1$$

如此，将积分区间分为$N=2^i$等份后，则有

$$S_{2^i}=\frac{4}{4-1}T_{2^{(i+1)}}-\frac{4}{4-1}T_{2^i}\quad(i=0,1,2,\cdots)$$

应注意，当积分区间由 $N=2^i$ 等份再分半变为 $2N=2^{i+1}$ 等份时，只增加了 N 个新节点，还有 $N+1$ 个原有的老节点。这些老节点上的被积函数值已在上次计算 T_N 时求出，在计算 T_{2N} 时可直接应用，不必重复计算。实际上按照复化梯形公式有

$$\begin{aligned}T_{2N}&=\frac{b-a}{4N}\left[f(a)-f(b)+2\sum_{k=1}^{N}f\left(a+2k\frac{b-a}{2N}\right)+2\sum_{k=1}^{N}f\left(a+(2k-1)\frac{b-a}{2N}\right)\right]\\&=\frac{1}{2}\frac{b-a}{2N}\left[f(a)-f(b)+2\sum_{k=1}^{N}f\left(a+k\frac{b-a}{N}\right)\right]+\frac{b-a}{2N}\sum_{k=1}^{N}f\left(a+(2k-1)\frac{b-a}{2N}\right)\\&=\frac{1}{2}T_N+\frac{b-a}{2N}\sum_{k=1}^{N}f\left(a+(2k-1)\frac{b-a}{2N}\right)\end{aligned}$$

于是，可得递推公式

$$\begin{cases}T_1=\dfrac{b-a}{2}[f(a)+f(b)]\\[2ex]T_{2^i}=\dfrac{1}{2}T_{2^{(i-1)}}+\dfrac{b-a}{2^i}\displaystyle\sum_{k=1}^{2^{(i-1)}}f\left(a+(2k-1)\frac{b-a}{2^i}\right)\quad(i=1,2,\cdots)\end{cases}$$

例 7-4 利用变步长梯形公式计算积分 $\int_0^1\frac{\sin x}{x}\mathrm{d}x$。

解 $T_1=\frac{1}{2}[f(0)+f(1)]=\frac{1}{2}[1+0.841\,471\,0]=0.920\,735\,5$

$$T_2=\frac{1}{2}T_1+\frac{1}{2}f\left(\frac{1}{2}\right)=\frac{1}{2}(0.920\,735\,5+0.958\,851\,1)=0.939\,793\,3$$

进行一次加速计算

$$\begin{aligned}S_1&=\frac{1}{4-1}T_2-\frac{1}{4-1}T_1=\frac{4}{3}\times0.939\,793\,3-\frac{1}{3}\times0.920\,735\,5\\&=0.946\,145\,9\\T_4&=\frac{1}{2}T_2+\frac{1}{4}\left[f\left(\frac{1}{4}\right)+f\left(\frac{3}{4}\right)\right]\\&=\frac{1}{2}\times0.939\,793\,3+\frac{1}{4}(0.989\,615\,8+0.908\,851\,7)\\&=0.944\,513\,5\end{aligned}$$

再进行加速计算

$$\begin{aligned}S_2&=\frac{4}{3}T_4-\frac{1}{3}T_2=\frac{4}{3}\times0.944\,513\,5-\frac{1}{3}\times0.939\,793\,3\\&=0.946\,086\,9\\C_1&=\frac{4^2}{4^2-1}S_2-\frac{1}{4^2-1}S_1=\frac{16}{15}\times0.946\,086\,9-\frac{1}{15}\times0.946\,145\,9\end{aligned}$$

$=0.946\,083$

由此可见，利用变步长加速计算，由梯形公式算至 T_4，然后由加速计算 C_1，已接近准确值。

二、龙贝格加速公式

由前述可知，复化辛普生公式前后两次积分结果的线性组合便是精度更高、收敛速度更快的复化柯特斯公式的计算结果，即

$$C_N=\frac{4^2}{4^2-1}S_{2N}-\frac{1}{4^2-1}S_N$$

事实上，当 $N=1$ 时

$$S_1=\frac{1}{6}(b-a)\left[f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right]$$

$$S_2=\frac{1}{12}(b-a)\left[f(a)+4f\left(\frac{3a+b}{4}\right)+2f\left(\frac{a+b}{2}\right)+4f\left(\frac{a+3b}{4}\right)+f(b)\right]$$

$$I=\int_a^b f(x)\,\mathrm{d}x=\frac{4^2}{4^2-1}S_2-\frac{1}{4^2-1}S_1$$

$$=\frac{16}{15}\times\frac{b-a}{12}\left[f(a)+4f\left(\frac{3a+b}{4}\right)+2f\left(\frac{a+b}{2}\right)+4f\left(\frac{a+3b}{4}\right)+f(b)\right]-$$

$$\frac{1}{15}\times\frac{b-a}{6}\left[f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right]$$

$$=\frac{b-a}{90}\left[7f(a)+32f\left(\frac{3a+b}{4}\right)+12f\left(\frac{a+b}{2}\right)+32f\left(\frac{a+3b}{4}\right)+7f(b)\right]$$

$$=C_1$$

即

$$I=C_1=\frac{4^2}{4^2-1}S_2-\frac{1}{4^2-1}S_1$$

这样，将区间 $[a,b]$ 分为 $N=2^i(i=1,2,\cdots)$ 等份时，就有

$$C_{2^i}=\frac{4^2}{4^2-1}S_{2^{(i+1)}}-\frac{1}{4^2-1}S_{2^i}\quad(i=1,2,\cdots)$$

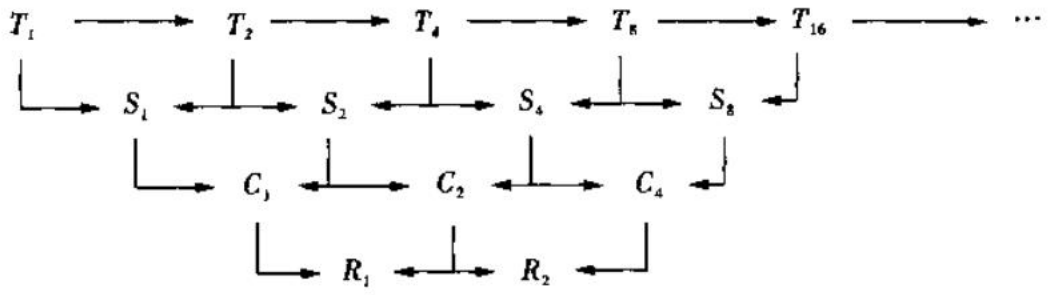
同理，对复化柯特斯公式的积分结果再进行线性组合，便使计算精度和收敛速度进一步提高。这种线性组合所得的求积公式叫做龙贝格加速公式。

$$R_{2^i}=\frac{4^3}{4^3-1}C_{2^{(i+1)}}-\frac{1}{4^3-1}C_{2^i}\quad(i=1,2,\cdots)$$

照此规律继续组合，求积公式的计算精度和收敛速度还可再提高。一般组合系数分别为 $\frac{4^m}{4^m-1}$ 和 $\frac{1}{4^m-1}(m=1,2,3,\cdots)$。但当 $m\geqslant 4$ 时，第一个系数接近于 1，而第二个系数的绝对值很小，则这样组合出来的求积公式与前一个公式的计算结果差别不大，因此，在

实际计算时，只做到龙贝格加速公式为止，即取 $m=3$。

从前述的推导过程可知，由精度较差的复化梯形公式的积分结果 $T_{2^i}(i=0,1,2,\cdots)$ 出发，可依次线性组合为精度越来越高的积分结果 $S_{2^i}, C_{2^i}, R_{2^i}, \cdots$，因此，连续 4 次复化梯形公式的积分结果直接线性组合为龙贝格加速公式。龙贝格求积公式的组合过程可写为：



事实上，对于 $i=0, N=2^0=1$，有

$$
\begin{aligned}
R_1 &= R_{2^0} = \frac{4^3}{4^3-1}C_{2^1} - \frac{1}{4^3-1}C_{2^0} \\
&= \frac{64}{63}\left(\frac{4^2}{4^2-1}S_{2^2} - \frac{1}{4^2-1}S_{2^1}\right) - \frac{1}{63}\left(\frac{4^2}{4^2-1}S_{2^1} - \frac{1}{4^2-1}S_{2^0}\right) \\
&= \frac{64}{63}\left[\frac{16}{15}\left(\frac{4}{4-1}T_{2^3} - \frac{1}{4-1}T_{2^2}\right) - \frac{1}{15}\left(\frac{4}{4-1}T_{2^2} - \frac{1}{4-1}T_{2^1}\right)\right] - \\
&\quad \frac{1}{63}\left[\frac{16}{15}\left(\frac{4}{4-1}T_{2^2} - \frac{1}{4-1}T_{2^1}\right) - \frac{1}{15}\left(\frac{4}{4-1}T_{2^1} - \frac{1}{4-1}T_{2^0}\right)\right] \\
&= \frac{1}{2\,835}(4\,096T_{2^3} - 1\,344T_{2^2} + 84T_{2^1} - T_{2^0})
\end{aligned}
$$

于是有

$$
R_{2^i} = \frac{1}{2\,835}(4\,096T_{2^{i+3}} - 1\,344T_{2^{i+2}} + 84T_{2^{i+1}} - T_{2^i}) \quad (i=0,1,2,\cdots)
$$

若再考虑到区间逐次分半的复化梯形公式积分值之间的递推关系，则龙贝格加速公式可写为如下格式

$$
\begin{cases}
T_1 = \dfrac{(a-b)}{2}[f(a)+f(b)] \\
h_T = (b-a)/2^i \\
T_{2^i} = \dfrac{1}{2}T_{2^{(i-1)}} + h_T\sum\limits_{k=1}^{2^{(i-1)}} f(a+2kh_T - h_T) \quad (i=1,2,\cdots) \\
R_{2^j} = \dfrac{1}{2\,835}(4\,096T_{2^{j+3}} - 1\,344T_{2^{j+2}} + 84T_{2^{j+1}} - T_{2^j}) \quad (j=0,1,\cdots,i-3)
\end{cases}
$$

显然，计算 1 次龙贝格积分值，需要连续 4 次复化梯形公式的积分结果。例如，对于 $j=0, R^{(0)}=R_{2^0}$，则需要 $T^{(i)}=T_{2^{(i-1)}}(i=1,2,3,4)$。于是起步时，先将区间分为 8 等份来求 $R^{(0)}$，即

$$\begin{cases} h=\dfrac{1}{8}(b-a) \\ T^{(1)}=T_{2^0}=T_1=4h[f(a)+f(b)] \\ T^{(2)}=T_{2^1}=T_2=\dfrac{1}{2}T^{(1)}+4hf(a+4h) \\ T^{(3)}=T_{2^2}=T_4=\dfrac{1}{2}T^{(2)}+2h[f(a+2h)+f(a+6h)] \\ T^{(4)}=T_{2^3}=T_8=\dfrac{1}{2}T^{(3)}+h\sum\limits_{k=1}^{4}f(a+2kh-h) \\ R^{(0)}=R_{2^0}=R_1=\dfrac{1}{2\,835}[4\,096T^{(4)}-1\,344T^{(3)}+84T^{(2)}-T^{(1)}] \end{cases}$$

然后再将区间逐次分半，应用变步长梯形公式的递推和龙贝格加速公式求 $R^{(j)}=R_{2^j}, j=1,2,\cdots$，即

$$\begin{cases} N=2^{(i-2)} \\ h=(b-a)/2^{(i-1)} \\ T^{(i)}=T_{2^{(i-1)}}=\dfrac{1}{2}T^{(i-1)}+h\sum\limits_{k=1}^{N}f(a+2kh-h) \quad (i=5,6,\cdots) \\ R^{(j)}=R_{2^j}=\dfrac{1}{2\,835}[4\,096T^{(j+4)}-1\,344T^{(j+3)}+84T^{(j+2)}-T^{(j+1)}] \quad (j=1,2,\cdots) \end{cases}$$

于是，龙贝格加速公式的计算步骤如下：

(1)取 $j=0, h=\dfrac{1}{8}(b-a), N=4$，求出 $R^{(0)}$。

(2)将区间逐次分半求 $R^{(j)}, j=1,2,\cdots$。

(3)若 $|[R^{(j)}-R^{(j-1)}]/R^{(j)}|\leqslant\varepsilon$，则终止计算，输出计算结果 $R^{(j)}$；否则，令 $N=2N$，$h=h/2$，回第(2)步。

例 7-5 利用龙贝格加速法计算积分 $\int_a^b \dfrac{\sin x}{x}\mathrm{d}x$。

解 将计算结果列于下表：

i	T_{2^i}	S_{2^i}	C_{2^i}	R_{2^i}
0	0.920 735 5	0.946 145 9	0.946 083 0	0.946 083 1
1	0.939 793 3	0.946 086 9	0.946 083 1	
2	0.944 513 5	0.946 083 3		
3	0.945 690 9			

显然，由计算出的 T_1, T_2, T_4 和 T_8，经过三步加速，即进行一次龙贝格计算，就具有相

当高的精度。若单独使用复化梯形公式，$T_{2^{10}}=0.946\ 083\ 1$，可见龙贝格加速公式具有很高的收敛速度。

第六节　高斯型求积公式

由前面的讨论知道，对于插值型求积公式

$$\int_a^b f(x)\mathrm{d}x = \sum_{k=1}^{n} A_k f(x_k) + R_n[f],\quad A_k = \int_a^b \frac{\omega(x)}{(x-x_k)\omega'(x_k)}\mathrm{d}x$$

它具有 $n+1$ 次代数精度，为此要提高牛顿—柯特斯求积公式的精度，就必须增加节点数。现提出这样的问题：若不限定节点等距分布，固定节点数，适当选取节点能否提高插值型求积公式的代数精度呢？答案是肯定的，高斯给出，只要合理选择 $n+1$ 个节点 x_k $(k=0,1,2,\cdots,n)$，就能使插值型求积公式代数精度达 $2n+1$ 次。称具有 $2n+1$ 次代数精度的内插型求积公式为高斯型求积公式，其节点 $x_k(k=0,1,2,\cdots,n)$ 称为高斯点。

这样，在固定节点数的条件下，要提高插值型求积公式代数精度的关键就是确定高斯点。

一、插值型求积公式的最高代数精度

为讨论插值型求积公式的最高代数精度，先介绍如下定理。

定理 1　在区间 $[-1,1]$ 上，节点 $x_k(k=0,1,2,\cdots,n)$ 是高斯点的充分必要条件是 $n+1$ 次多项式

$$\omega(x) = (x-x_0)(x-x_1)\cdots(x-x_n) = \prod_{k=0}^{n}(x-x_k)$$

与一切次数不超过 n 次的多项式 $q(x)$ 都正交，即

$$\int_{-1}^{1} q(x)\omega(x)\mathrm{d}x = 0$$

证明　必要性，设 $x_0,x_1,x_2,\cdots,x_n$ 是高斯点。则求积公式

$$\int_{-1}^{1} f(x)\mathrm{d}x = \sum_{k=0}^{n} A_k f(x_k)$$

对于 $2n+1$ 次多项式都精确成立。

而 $q(x)$ 的次数不超过 n，$\omega(x)$ 为 $n+1$ 次多项式，那么，$q(x)\omega(x)$ 的次数不超过 $2n+1$，且 $\omega(x_k)=0$，于是有

$$\int_{-1}^{1} q(x)\omega(x)\mathrm{d}x = \sum_{k=0}^{n} A_k q(x_k)\omega(x_k) = 0$$

即 $q(x)$ 与 $\omega(x)$ 正交。

充分性，设 $\omega(x)$ 与一切不超过 n 次的多项式正交，若 $f(x)$ 是次数不超过 $2n+1$ 次的多项式，用 $\omega(x)$ 除 $f(x)$ 得

$$f(x) = q(x)\omega(x) + \gamma(x)$$

其中，$q(x)$和$\gamma(x)$的次数不超过n，从而

$$\int_{-1}^{1} f(x)\mathrm{d}x = \int_{-1}^{1} q(x)\omega(x)\mathrm{d}x + \int_{-1}^{1} \gamma(x)\mathrm{d}x = \int_{-1}^{1} \gamma(x)\mathrm{d}x$$

于是

$$\int_{1}^{1} \gamma(x)\mathrm{d}x = \sum_{k=0}^{n} A_k \gamma(x_k)$$

精确成立。而$f(x_k) = \gamma(x_k)$，所以

$$\int_{-1}^{1} f(x)\mathrm{d}x = \sum_{k=0}^{n} A_k \gamma(x_k)$$

也即$x_k(k = 0,1,2,\cdots,n)$是高斯点。

定理2 在区间$[-1,1]$内，积分$\int_{-1}^{1} f(x)\mathrm{d}x$的高斯型求积公式

$$\int_{-1}^{1} f(x)\mathrm{d}x = \sum_{k=0}^{n} A_k f(x_k), \quad A_k = \int_{-1}^{1} \frac{\omega(x)}{(x - x_k)\omega'(x_k)}\mathrm{d}x$$

的最高代数精度为$2n+1$次。

证明 由定理1的充分性已经证明了高斯型求积公式具有$2n+1$次代数精度。现证明$2n+1$为其最高代数精度。

若取

$$f(x) = (x - x_0)^2 (x - x_1)^2 \cdots (x - x_n)^2$$

则

$$\int_{-1}^{1} f(x)\mathrm{d}x > 0$$

而

$$\sum_{k=0}^{n} A_k f(x_k) = 0$$

则

$$\int_{-1}^{1} f(x)\mathrm{d}x \neq \sum_{k=0}^{n} A_k f(x_k)$$

即高斯型求积公式不具有$2n+2$次代数精度，$2n+1$为其最高代数精度。

由上述两定理可见，要得到$2n+1$次代数精度的求积公式，就要找到高斯点，即求出$n+1$次正交多项式$\omega(x) = \prod_{k=0}^{n}(x - x_k)$的$n+1$个不相等的实根，而勒让德多项式就具有这种性质，也就是说，用$n+1$次勒让德多项式的$n+1$个不相等的实根作为节点，构造出的插值型求积公式具有$2n+1$次代数精度。

二、勒让德多项式

1. 勒让德多项式的定义

n 阶勒让德多项式是区间$[-1,1]$上的一个 n 次多项式，其罗德利格斯表达式为

$$L_n(x)=\frac{1}{2^n n!}\frac{\mathrm{d}^n}{\mathrm{d}x^n}[(x^2-1)^n] \quad (n=0,1,2,\cdots)$$

它是以 n 阶导数的形式表示的，由此可以将前几个勒让德多项式写出来

$$L_0=1$$

$$L_1(x)=x$$

$$L_2(x)=\frac{1}{2}(3x^2-1)$$

$$L_3(x)=\frac{1}{2}(5x^2-3x)$$

$$L_4(x)=\frac{1}{8}(35x^4-30x^2+3)$$

$$L_5(x)=\frac{1}{8}(63x^5-70x^3+15x)$$

……

从而可得勒让德多项式的一般形式为

$$L_{2n}(x)=\frac{1}{2^{2n}}\sum_{k=0}^{n}\frac{(-1)^k(4n-2k)!}{k!(2n-k)!(2n-2k)!}x^{2n-2k}$$

$$L_{2n+1}(x)=\frac{1}{2^{2n+1}}\sum_{k=0}^{n}\frac{(-1)^k(4n+2-2k)!}{k!(2n+1-k)!(2n+1-2k)!}x^{2n+1-2k}$$

于是有

$$L_{2n+1}(0)=0,\quad L_{2n}(0)=\frac{(-1)^n\cdot(2n)!}{2^{2n}\cdot(n!)^2}$$

2. 勒让德多项式的性质

(1) $L_n(-x)=(-1)^nL_n(x)$，即 n 为偶数时，$L_n(x)$ 为偶函数；n 为奇数时，$L_n(x)$ 为奇函数。

(2) 正交性。勒让德多项式函数$\{L_n(x)\}(n=0,1,2,\cdots)$在区间$[-1,1]$上构成正交函数族，即

$$\int_{-1}^{1}L_m(x)L_n(x)\mathrm{d}x=\begin{cases}0 & (m\neq n)\\ \dfrac{2}{2n+1} & (m=n)\end{cases}$$

证明 考虑积分

$$(L_m,L_n)=\int_{-1}^{1}L_m(x)L_n(x)\mathrm{d}x$$

$$=\frac{1}{2^{(m+n)}\cdot m!\cdot n!}\int_{-1}^{1}\frac{\mathrm{d}^n}{\mathrm{d}x^n}[(x^2-1)^n]\frac{\mathrm{d}^m}{\mathrm{d}x^m}[(x^2-1)^m]$$

当 $m\neq n$（设 $m<n$）时，连续使用 n 次分步积分有

$$(L_m,L_n)=\frac{1}{2^{(m+n)}\cdot m!\cdot n!}\left\{\frac{\mathrm{d}^m}{\mathrm{d}x^m}[(x^2-1)^m]\frac{\mathrm{d}^{n-1}}{\mathrm{d}x^{n-1}}[(x^2-1)^n]\Big|_{-1}^{1}-\right.$$
$$\left.\int_{-1}^{1}\frac{\mathrm{d}^{m+1}}{\mathrm{d}x^{m+1}}[(x^2-1)^m]\frac{\mathrm{d}^{n-1}}{\mathrm{d}x^{n-1}}[(x^2-1)^n]\mathrm{d}x\right\}$$
$$=\cdots\cdots$$
$$=(-1)^n\frac{1}{2^{(m+n)}\cdot m!\cdot n!}\int_{-1}^{1}\frac{\mathrm{d}^{m+n}}{\mathrm{d}x^{m+n}}[(x^2-1)^m](x^2-1)^n\mathrm{d}x$$

由于 $m<n$，则 $\frac{\mathrm{d}^{m+n}}{\mathrm{d}x^{m+n}}[(x^2-1)^m]=0$，于是

$$(L_m,L_n)=\int_{-1}^{1}L_m(x)L_n(x)\mathrm{d}x=0$$

当 $m=n$ 时

$$(L_m,L_n)=\left(\frac{1}{2^n n!}\right)^2\int_{-1}^{1}\frac{\mathrm{d}^n}{\mathrm{d}x^n}[(x^2-1)^n]\frac{\mathrm{d}^n}{\mathrm{d}x^n}[(x^2-1)^n]\mathrm{d}x$$

也连续使用 n 次分步积分得

$$(L_m,L_n)=(-1)^n\left(\frac{1}{2^n n!}\right)^2\int_{-1}^{1}(2n)!(x^2-1)^n\mathrm{d}x$$
$$=(-1)^n\left(\frac{1}{2^n n!}\right)^2 2(2n)!\int_{-1}^{1}(2n)!(x^2-1)^n\mathrm{d}x$$

令 $x=\sin t,\mathrm{d}x=\cos t\mathrm{d}t$，当 $x=0$ 时，$t=0$；当 $x=1$ 时，$t=\frac{\pi}{2}$，则

$$(L_m,L_n)=\frac{(2n)!}{2^{2n-1}(n!)^2}\int_0^{\frac{\pi}{2}}\cos^{2n+1}t\mathrm{d}t=\frac{2}{2n+1}$$

则

$$(L_m,L_n)=\int_{-1}^{1}L_m(x)L_n(x)\mathrm{d}x=\begin{cases}0 & (m\neq n)\\ \dfrac{2}{2n+1} & (m=n)\end{cases}$$

即 $\{L_n(x)\}$ 是 $[-1,1]$ 上的正交函数族，$L_n(x)$ 为 n 次正交多项式。

三、高斯—勒让德求积公式

若在区间 $[-1,1]$ 内取

$$\omega(x)=L_{n+1}(x)=\frac{1}{2^{(2n+1)}(n+1)!}\frac{\mathrm{d}^{n+1}}{\mathrm{d}x^{n+1}}[(x^2-1)^{n+1}]\quad(n=0,1,2,\cdots,n)$$

于是勒让德多项式 L_{n+1} 的 $n+1$ 个实根便为高斯点 $x_k(k=0,1,2,\cdots,n)$，即

$$L_{n+1}(x_k)=0\quad(k=0,1,2,\cdots,n)$$

则求积公式

$$\int_{-1}^{1}f(x)\mathrm{d}x=\sum_{k=0}^{n}A_k f(x_k),\quad A_k=\int_{-1}^{1}\frac{L_{n+1}(x)}{(x-x_k)L'_{n+1}(x_k)}\mathrm{d}x$$

称为高斯—勒让德求积公式。A_k 为高斯型求积公式的求积系数，简称高斯求积系数。

高斯—勒让德求积公式在 $[-1,1]$ 内具有 $2n+1$ 次代数精度，高斯求积系数 A_k 可由如下定理求得。

定理 高斯求积系数 A_k 都是正的，并且

$$A_k = \int_{-1}^{1} \frac{L_{n+1}(x)}{(x-x_k)L'_{n+1}(x_k)}\mathrm{d}x = \frac{2}{(1-x_k^2)[L'_{n+1}(x_k)]^2} \quad (k=0,1,2,\cdots,n)$$

证明 因高斯型求积公式具有 $2n+1$ 次代数精度，而 $\dfrac{L_{n+1}(x)}{(x-x_k)}L'_{n+1}(x)$ 是 $2n$ 次多项式，则

$$\int_{-1}^{1} \frac{L_{n+1}(x)}{(x-x_k)}L'_{n+1}(x)\mathrm{d}x = \sum_{j=0}^{n} A_j \frac{L_{n+1}(x_j)}{(x_j-x_k)}L'_{n+1}(x_j)$$

但

$$\frac{L_{n+1}(x_j)}{(x_j-x_k)} = \begin{cases} 0 & (j\neq k) \\ L'_{n+1}(x_k) & (j=k) \end{cases}$$

于是

$$\int_{-1}^{1} \frac{L_{n+1}(x)}{(x-x_k)}L'_{n+1}(x)\mathrm{d}x = A_k[L'_{n+1}(x_k)]^2$$

又由分步积分法可得

$$\begin{aligned}\int_{-1}^{1} \frac{L_{n+1}(x)}{(x-x_k)}L'_{n+1}(x)\mathrm{d}x &= \int_{-1}^{1} \frac{L_{n+1}(x)}{(x-x_k)}\mathrm{d}L_{n+1}(x) \\ &= \frac{L_{n+1}^2(x)}{(x-x_k)}\bigg|_{-1}^{1} - \int_{-1}^{1} L_{n+1}(x)\left[\frac{L_{n+1}(x)}{(x-x_k)}\right]'\mathrm{d}x\end{aligned}$$

由于 $L_{n+1}(x_k)=0$，$L_{n+1}(x)\left[\dfrac{L_{n+1}(x)}{(x-x_k)}\right]'$ 为 $2n$ 次多项式，则

$$\int_{-1}^{1} L_{n+1}(x)\left[\frac{L_{n+1}(x)}{(x-x_k)}\right]'\mathrm{d}x = 0$$

于是

$$\int_{-1}^{1} \frac{L_{n+1}(x)}{(x-x_k)}L'_{n+1}(x)\mathrm{d}x = \frac{L_{n+1}^2(1)}{(1-x_k)} + \frac{L_{n+1}^2(-1)}{(1+x_k)}$$

而

$$L_{n+1}(-1) = (-1)^{n+1}L_{n+1}(1), \quad L_{n+1}(1)=1$$

$$\int_{-1}^{1} \frac{L'_{n+1}(x)}{(x-x_k)}L_{n+1}(x)\mathrm{d}x = \frac{2}{1-x_k^2} = A_k[L'_{n+1}(x_k)]^2$$

故

$$A_k = \frac{2}{(1-x_k^2)[L'_{n+1}(x_k)]^2}$$

因 $x_k \in [-1,1]$，则 $A_k > 0$。

由此式即可算出不同 n 值下的高斯求积系数 A_k，例如，当 $n=0$ 时，

$$L_1(x)=\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}[(x^2-1)]=x,\qquad L'_1(x)=1$$

勒让德多项式为一次，它只有 1 个零点，即 $x=0$，且 $L'_1(0)=1$，所以

$$A_0=\frac{2}{(1-0)\times 1}=2$$

因而，求积公式为

$$\int_{-1}^{1}f(x)\mathrm{d}x=\sum_{k=0}^{n}A_kf(x_k)=2f(x_0)$$

当 $n=1$ 时

$$L_2(x)=\frac{1}{2}(3x^2-1),\quad L'_2(x)=3x$$

勒让德多项式为二次多项式，其两个零点分别为

$$x_0=-\frac{1}{3}\sqrt{3},\quad x_1=\frac{1}{3}\sqrt{3}$$

$$A_0=\frac{2}{(1-x_0^2)[L'_2(x_0)]^2}=\frac{2}{(1-\frac{1}{3})(-\sqrt{3})^2}=1$$

$$A_1=\frac{2}{(1-x_1^2)[L'_2(x_1)]^2}=\frac{2}{(1-\frac{1}{3})(\sqrt{3})^2}=1$$

于是，求积公式为

$$\int_{-1}^{1}f(x)\mathrm{d}x=\sum_{k=0}^{1}A_kf(x_k)=f\left(-\frac{\sqrt{3}}{3}\right)+\left(\frac{\sqrt{3}}{3}\right)$$

实际应用时，由于 x_k，A_k 可事先算出，则可直接代入求积公式

$$\int_{-1}^{1}f(x)\mathrm{d}x=\sum_{k=0}^{n}A_kf(x_k)$$

中进行计算。表 7-2 中列出了部分高斯点及其相应的求积系数。

可以推得高斯型求积公式的余项

$$R_n[f]=\frac{2^{2n+1}}{2n+1}\frac{(n!)^4}{[(2n)!]^3}f^{(2n)}(\eta)\quad(-1<\eta<1)$$

当 $n=2$ 时，高斯型求积公式的代数精度为 5 次，余项 $R_2(f)=\frac{1}{135}f^{(4)}(\eta)$。可见高斯型求积公式的精度相当高。

表 7-2

$n+1$	1	2	3	4	5
x_k	0	±0.577 350 3	0 ±0.774 596 7	±0.339 981 0 ±0.861 136 3	±0.538 469 3 0 ±0.906 179 9
A_k	2.000 000	1.000 000	0.888 888 9 0.555 555 6	0.652 145 2 0.347 854 8	0.478 628 7 0.568 888 9 0.236 926 9

然而，上述的高斯—勒让德求积公式只适合于在区间$[-1,1]$上的积分，工程计算上通常遇到的是在任意区间$[a,b]$上的积分，这时，需进行适当的变换，将$x\in[a,b]$映射到$t\in[-1,1]$上，即作变换

$$x=\frac{a+b}{2}+\frac{b-a}{2}t,\quad \mathrm{d}x=\frac{b-a}{2}\mathrm{d}t$$

则

$$\begin{aligned}\int_b^a f(x)\,\mathrm{d}x &= \frac{b-a}{2}\int_{-1}^{1} f\left(\frac{a+b}{2}+\frac{b-a}{2}t\right)\mathrm{d}t\\ &=\frac{b-a}{2}\sum_{k=0}^{n}A_k\cdot f\left(\frac{a+b}{2}+\frac{b-a}{2}t_k\right)\\ &=\frac{b-a}{2}\sum_{k=0}^{n}A_k f(x_k)\end{aligned}$$

其中，$x_k=\frac{a+b}{2}+\frac{b-a}{2}t_k\quad(k=0,1,2,\cdots,n)$。

由于A_k与$f(x_k)=f\left(\frac{a+b}{2}+\frac{b-a}{2}t_k\right)$无关，它只与$x_k=\frac{a+b}{2}+\frac{b-a}{2}t_k$有关，因此上述的各$A_k$在$[a,b]$上不变。只要求得高斯点$t_k(k=0,1,2,\cdots,n)$所对应的$x_k(k=0,1,2,\cdots,n)$，则可求得函数值$f(x_k)$。从而即可由式

$$\int_b^a f(x)\,\mathrm{d}x=\frac{b-a}{2}\sum_{k=0}^{n}A_k f(x_k)$$

计算积分结果。

例 7-6 用高斯型求积公式计算$\int_1^2\frac{1}{x}\mathrm{d}x$。

解 由于$a=1,b=2$，则$x=\frac{3}{2}+\frac{1}{2}t$，

$$\int_1^2\frac{1}{x}\mathrm{d}x=\frac{1}{2}\int_{-1}^{1}\frac{\mathrm{d}t}{\frac{3}{2}+\frac{1}{2}t}=\int_{-1}^{1}\frac{1}{3+t}\mathrm{d}t$$

$$=0.555\,555\,6\times\frac{1}{3-0.774\,596\,7}+\frac{0.888\,888\,9}{3}+\frac{0.555\,555\,6}{3+0.774\,596\,7}$$

$$=0.693\,12$$

而精确解

$$\int_1^2\frac{1}{x}\mathrm{d}x=\ln x\Big|_1^2=\ln 2\approx 0.693\,15$$

若将积分区间$[a,b]$分为m个小区间$[a_i,b_i](i=1,2,\cdots,m)$，在每个小区间上施以高斯型求积公式，便构成了复化高斯型求积公式

$$\int_a^b f(x)\mathrm{d}x=\sum_{i=1}^m\int_{a_i}^{b_i}f(x)\mathrm{d}x\approx\frac{h}{2}\sum_{i=1}^m\sum_{k=0}^n A_k f(x_k)$$

其中

$$h=b_i-a_i=(b-a)/m$$

$$x_k=\frac{b_i+a_i}{2}+\frac{b_i-a_i}{2}t_k=\frac{1}{2}[2a+h(2i-1)+ht_k]\quad(i=1,2,\cdots,m;k=0,1,2,\cdots,n)$$

习　题

1. 已知$f(x)=\dfrac{1}{(1+x)^2}$，用三点和五点公式求$x=1.1,1.2$及1.3处的导数值，并估计其误差。数据如下：

x	1.0	1.1	1.2	1.3	1.4
$f(x)$	0.250 0	0.226 8	0.206 6	0.189 8	0.173 6

2. 试分别用复化梯形公式和复化辛普生公式计算下列积分，并比较结果。

(1)$\int_0^1\dfrac{x}{4+x^2}\mathrm{d}x,N=8$；　　(2)$\int_0^1\sqrt{x}\mathrm{d}x,N=4$；

(3)$\int_0^1 e^{-x}\mathrm{d}x,N=6$；　　(4)$\int_1^2\dfrac{x}{1+\ln x}\mathrm{d}x,N=4$。

3. 用复化柯特斯公式计算积分$\int_0^{\pi}\ln(5-4\cos x)\mathrm{d}x$，要求精度$\varepsilon=10^{-5}$。

4. 已知某函数在等距节点上的值为：

x	0	1	2	3	4	5	6	7	8
$f(x)$	0	0.568 7	0.790 9	0.574 3	0.135 0	−0.185 2	−0.180 2	0.081 1	0.291 7

试用适宜的方法求 $\int_0^8 f(x)\,dx$ 的值，要求 $\varepsilon = 10^{-5}$。

5. 试用龙贝格加速法求下列积分，要求精度 $\varepsilon = 10^{-5}$。

(1) $\int_0^1 \frac{4}{1+x^2}dx$；　　(2) $\frac{2}{\sqrt{\pi}}\int_0^1 e^{-x^2}dx$；

(3) $\int_0^1 \frac{\ln(1+x)}{1+x^2}dx$；　　(4) $\int_0^1 \frac{1}{1+x^3}dx$。

6. 试用高斯型求积公式计算下列积分，精度要求 $\varepsilon = 10^{-5}$。

(1) $\int_0^1 \frac{\sqrt{x}}{(1+x)^2}dx$；　　(2) $\int_0^1 \frac{1}{1+x^2}dx$；

(3) $\int_0^1 \sqrt{1+2x}dx$；　　(4) $\int_0^{\pi} x\sin x dx$。

第八章　常微分方程的数值解法

微分方程是用来描述现象的一种重要手段，对它们的求解在工业过程分析、设计中是必不可少的工作。化工过程中遇到的微分方程问题通常分为两类：一类是化工过程的动态分析，即研究过程参量随时间的变化规律，如间歇反应器内不同时刻反应物和产物的浓度，非稳态的传热和传质的分析等；另一类是计算稳态过程中参量的一维分布，例如，等温条件下，管式反应器中反应物或产物沿管长的浓度分布等。通常第一类中遇到的主要是初值问题，即已知自变量初始点处的函数值以及导数值的情况，第二类遇到的主要是边值问题，它是指给出自变量边界处的函数值或导数值的情况，两者只是给定的已知条件的不同，在求解方法上无原则差别。

对于常微分方程的初值问题一般可表示为

$$\begin{cases}\dfrac{dy}{dx}=f(x,y), & x\in[x_0,x_n]\\ y|_{x=x_0}=y_0\end{cases}$$

其中，$f(x,y)$ 为已知函数；y_0 为已知初值；$[x_0,x_n]$ 为函数 $y(x)$ 的定义区间。

对于常微分方程的边值问题一般可表示为

$$\begin{cases}\dfrac{d^2y}{dx^2}=f\left(x,y,\dfrac{dy}{dx}\right), & x\in[x_0,x_n]\\ y|_{x=x_0}=y_0\\ y|_{x=x_n}=y_n\end{cases}$$

其中，$f\left(x,y,\dfrac{dy}{dx}\right)$ 为已知函数；y_0,y_n 为已知边界条件。当然，边界条件也可以按照实际问题以其他方式给出。该式称为两点边值问题。

初值问题的数值解法是将微分方程离散化，即将区间 $x\in[x_0,x_n]$ 等距分割：

$$h=x_{i+1}-x_i=(x_n-x_0)/n$$

其中，h 称为步长。则各分点为

$$x_i=x_0+ih \quad (i=0,1,2,\cdots,n)$$

然后，利用已知条件$y|_{x=x_0}=y_0$，由离散方程求取函数$y(x)$在各分点上的近似值$y_i\approx y(x_i)$ $(i=1,2,\cdots,n)$，这种方法称为步进法。显然，离散化的方法是数值解法的关键。

至于边值问题，一般是将其转化为初值问题迭代求解，而对于线性常微分方程可应用差分法求解，在求解化工计算中的边值问题时，也常采用正交配置法。

第一节　解初值问题的尤拉法

尤拉法体现了初值问题数值解法的基本思想，是其他方法发展的基础。它的基本思路是应用均差代替导数，使微分方程离散化。采用不同类型的差商，可构成不同类型的差分方程，从而求得的近似解也有所不同。

一、显式尤拉公式

对于常微分方程的初值问题

$$\begin{cases}\dfrac{dy}{dx}=f(x,y), & x\in[x_0,x_n]\\ y|_{x=x_0}=y_0\end{cases}$$

现用一阶向前差商近似地代替任意点x_i处的一阶导数

$$\left.\frac{dy}{dx}\right|_{x=x_i}\approx\frac{y(x_{i+1})-y(x_i)}{x_{i+1}-x_i}=\frac{y(x_{i+1})-y(x_i)}{h}$$

则

$$\frac{y(x_{i+1})-y(x_i)}{h}\approx f(x_i,y_i)$$

$$y(x_{i+1})=y(x_i)+hf(x_i,y_i)\quad(i=0,1,2,\cdots,n-1)$$

若令$y_{i+1}\approx y(x_{i+1})$，$y_i\approx y(x_i)$，则有

$$\begin{cases}y_{i+1}=y_i+hf(x_i,y_i)\\ x_i=x_0+ih\end{cases}\qquad(i=0,1,2,\cdots,n)$$

该式即为显式尤拉公式。由(x_0,y_0)出发，可逐步求得近似数值解(x_i,y_i) $(i=0,1,2,\cdots,n)$。这样便将微分方程的求解问题转化为了代数方程组的求解问题。

显式尤拉公式的几何意义（如图8－1所示）是利用连续折线$y_{i+1}=y_i+hf(x_i,y_i)$ $(i=0,1,2,\cdots,n-1)$近似表示曲线$y=y(x)$。两个方程的初始点均为(x_0,y_0)，当$n\to\infty$时，$h\to0$，曲线与折线重合。因此，显式尤拉公式可收敛于常微分方程初值问题的解。

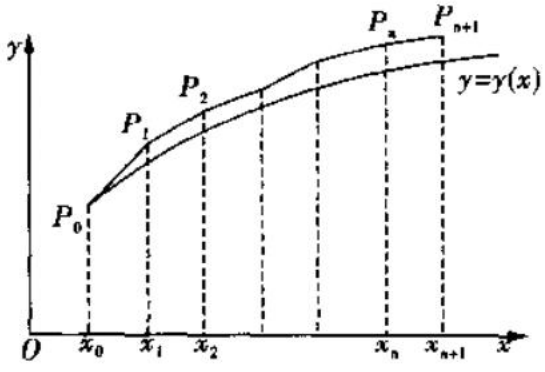


图8－1

现在讨论显式尤拉公式的计算误差。若假定第i步及其以前各步的计算无误差，即$y_i=y(x_i)$，

则第 $i+1$ 步的误差称为局部截断误差。为此将函数 $y(x)$ 在 x_i 处作泰勒展开：

$$y(x_{i+1})=y_i+(x_{i+1}-x_i)\left.\frac{dy}{dx}\right|_{x_i}+\frac{1}{2}(x_{i+1}-x_i)^2\left.\frac{d^2y}{dx^2}\right|_{\xi_i}\quad(x_i<\xi_i<x_{i+1})$$

而显式尤拉公式

$$y_{i+1}=y_i+h\left.\frac{dy}{dx}\right|_{x_i}$$

于是

$$y_{i+1}-y(x_{i+1})=-\frac{1}{2}h^2f'(\xi_i,y(\xi_i))=o(h^2)$$

即显式尤拉公式的局部截断误差为 $o(h^2)$，它是步长 h 的二阶小量。实际上，除了 $y(x_0)=y_0$ 之外，从 y_1 开始便有误差，而且这些误差在以后每步计算中均会累积。可以证明，显式尤拉公式的总误差为 $o(h)$，即与步长为同阶小量，于是显式尤拉公式也称为一阶法。

二、隐式尤拉公式

若应用一阶向后差商近似代替点 x_{i+1} 处的导数，则有

$$y(x_{i+1})\approx y(x_i)+hf(x_{i+1},y(x_{i+1}))$$

因而有

$$\begin{cases}y_{i+1}=y_i+hf(x_{i+1},y_{i+1})\\x_{i+1}=x_0+(i+1)h\end{cases}\quad(i=0,1,2,\cdots,n-1)$$

此式为隐式形式，因而称为隐式尤拉公式，需用迭代法求解。其几何意义是用另一条连续折线 $y_{i+1}=y_i+hf(x_{i+1},y_{i+1})(i=0,1,2,\cdots,n-1)$ 近似表示曲线 $y=(x)$。

隐式尤拉公式的计算格式可写为

$$\begin{cases}y_{i+1}^{(0)}=y_i+hf(x_i,y_i)\\y_{i+1}^{(k+1)}=hf(x_{i+1},y_{i+1}^{(k)})\quad(i=0,1,2,\cdots)\\x_{i+1}=x_i+(i+1)h\quad(i=0,1,2,\cdots,n-1)\end{cases}$$

即由显式尤拉公式提供初始值 $y_{i+1}^{(0)}(i=0,1,2,\cdots,n-1)$，再由隐式尤拉公式迭代求取近似解。当 $|y_{i+1}^{(k+1)}-y_{i+1}^{(k)}|\leqslant\varepsilon$ 时，则取 $y(x_{i+1})\approx y_{i+1}^{(k+1)}(i=0,1,2,\cdots,n-1)$。

为分析隐式尤拉公式的误差，对 $f(x_{i+1},y_{i+1})$ 应用中值定理有

$$f(x_{i+1},y_{i+1})=f(x_{i+1},y(x_{i+1}))+f_y'(x_{i+1},\eta)[y_{i+1}-y(x_{i+1})]$$

其中，$\eta\in[y_{i+1},y(x_{i+1})]$，$f_y'(x_{i+1},\eta)=\left.\frac{\partial f}{\partial y}\right|_{y=\eta}$。

因此，隐式尤拉公式可写为

$$y_{i+1}=y_i+hf(x_{i+1},y(x_{i+1}))+hf_y'(x_{i+1},\eta)[y_{i+1}-y(x_{i+1})]$$

而函数 $y(x)$ 在 x_i 处的三阶泰勒展开式为

$$y(x_{i+1}) = y_i + hf(x_i, y_i) + \frac{1}{2}h^2 f'(x_i, y_i) + \frac{h^3}{3!} f''(\xi_i, y(\xi_i))$$

$$x_i \leqslant \xi_i \leqslant x_{i+1}$$

于是

$$y_{i+1} - y(x_{i+1}) = h[f(x_{i+1}, y(x_{i+1})) - f(x_i, y_i)] + hf'_y(x_{i+1}, \eta)[y_{i+1} - y(x_{i+1})] - \frac{h^2}{2} f'(x_i, y_i) - \frac{h^3}{3!} f''(\xi_i, y(\xi_i))$$

其中 $f(x_{i+1}, y(x_{i+1})) = y'(x_{i+1})$。对其在 x_i 处作二阶泰勒展开：

$$f(x_{i+1}, y(x_{i+1})) = f(x_i, y_i) + hf'(x_i, y_i) + \frac{h^2}{2} f''(\eta_i, y(\eta_i)) \quad (x_i < \eta_i < x_{i+1})$$

则

$$y_{i+1} - y(x_{i+1}) = h^2 f'(x_i, y_i) + \frac{1}{2}h^3 f''(\eta_i, y(\eta_i)) + hf'_y(x_{i+1}, \eta)[y_{i+1} - y(x_{i+1})] - \frac{1}{2}h^2 f'(x_i, y_i) - \frac{1}{3!}h^3 f''(\xi_i, y(\xi_i))$$

$$= \frac{1}{2}h^2 f'(x_i, y_i) + hf'_y(x_{i+1}, \eta)[y_{i+1} - y(x_{i+1})] + \frac{1}{3!}h^3 [3f''(\eta_i, y(\eta_i)) - f''(\xi_i, y(\xi_i))]$$

当 $f''(x, y)$ 在 $[x_i, x_{i+1}]$ 上变化不大时，近似地取

$$f''(\eta_i, y(\eta_i)) \approx f''(\xi_i, y(\xi_i))$$

于是

$$y_{i+1} - y(x_{i+1}) = \frac{1}{1 - hf'_y(x_{i+1}, \eta)} \left[\frac{h^2}{2} f'(x_i, y_i) + \frac{h^3}{3} f''(\eta_i, y(\eta_i)) \right]$$

当 $h \to 0$ 时，$|hf'_y(x_{i+1}, \eta)| \ll 1$，故有

$$y_{i+1} - y(x_{i+1}) = \frac{1}{2}h^2 f'(x_i, y_i) + \frac{1}{3}h^3 f''(\eta_i, y(\eta_i)) = o(h^2)$$

故隐式尤拉公式的局部截断误差亦为 $o(h^2)$，总误差为 $o(h)$，即也为一阶方法。

三、改进的尤拉公式

显式尤拉公式的误差为

$$y_{i+1} - y(x_{i+1}) = -\frac{1}{2}h^2 f'(x_i, y_i) - \frac{1}{3!}h^3 f''(\xi_i, y(\xi_i)) \quad (x_i \leqslant \xi_i \leqslant x_{i+1})$$

隐式尤拉公式的误差为

$$y_{i+1} - y(x_{i+1}) = \frac{1}{2}h^2 f'(x_i, y_i) + \frac{1}{3}h^3 f''(\eta_i, y(\eta_i)) \quad (x_i \leqslant \eta_i \leqslant x_{i+1})$$

以上两式相加后整理得

$$y_{i+1}-y(x_{i+1})=\frac{h^3}{2\times 3!}f''(\eta_i,y(\eta_i))=o(h^3)$$

这里仍然近似取$f''(\eta_i,y(\eta_i))\approx f''(\xi_i,y(\xi_i))$。

于是可得到

$$\begin{cases}y_{i+1}=y_i+\dfrac{1}{2}h[f(x_i,y_i)+f(x_{i+1},y_{i+1})] \\ x_{i+1}=x_0+(i+1)h\end{cases}\quad(i=0,1,2,\cdots,n-1)$$

此即为改进尤拉公式，其局部截断误差为$o(h^3)$，总误差为$o(h^2)$，是一种二阶方法。

由于改进尤拉公式是隐式形式，需采用迭代法求解，即

$$\begin{cases}y_{i+1}^{(0)}=y_i+hf(x_i,y_i) \\ y_{i+1}^{(k+1)}=y_i+\dfrac{1}{2}h[f(x_i,y_i)+f(x_{i+1},y_{i+1}^{(k)})] \\ x_{i+1}=x_0+(i+1)h\end{cases}\quad(k=0,1,2,\cdots;i=0,1,2,\cdots,n-1)$$

由显式尤拉公式提供初始值。当$|y_{i+1}^{(k+1)}-y_{i+1}^{(k)}|\leqslant\varepsilon$时，取$y(x_{i+1})\approx y_{i+1}^{(k+1)}$($i=0,1,2,\cdots,n-1$)。

在一般情况下，采用预报—校正格式，即进行一次预报和一次校正，于是

$$\begin{cases}y_{i+1}^{(0)}=y_i+hf(x_i,y_i) \\ y_{i+1}=y_i+\dfrac{1}{2}h[f(x_i,y_i)+f(x_{i+1},y_{i+1}^{(0)})] \\ x_{i+1}=x_0+(i+1)h\end{cases}\quad(i=0,1,2,\cdots,n-1)$$

利用显式尤拉公式求出$y_{i+1}^{(0)}$为预报值，其误差为$o(h)$，再由改进尤拉式公式求y_{i+1}为校正值，其误差为$o(h^2)$，即利用改进尤拉公式计算精度提高了一阶，但计算函数的次数也增加一次。精度的提高是由增大计算工作量来获得的。

改进尤拉公式的预校式通常也称为梯形公式。其几何意义是在$[x_i,x_{i+1}]$内取两个端点的斜率平均值作为斜率的直线近似表示曲线。由图8-2可见，当已知点$A(x_i,y_i)$求点$B(x_{i+1},y_{i+1})$的近似值时，若用显式尤拉公式，则是通过点A作切线AC，C点的纵坐标y_{i+1}为$y(x_{i+1})$的近似值，具有负偏差；若采用隐式尤拉公式，则应过A点作过点$B(x_{i+1},y(x_{i+1}))$的切线BE的平行线AD，D点的纵坐标y_{i+1}为$y(x_{i+1})$的近似值，具有正偏差；若采用梯形公式，则是将C和D两点的值加以平均得F点，将其纵坐标y_{i+1}作为$y(x_{i+1})$的近似值。显然，F点比C点和D点更靠近B点，因而梯形公式具有较高的精度。为了便于计算，通常将梯形公式写为

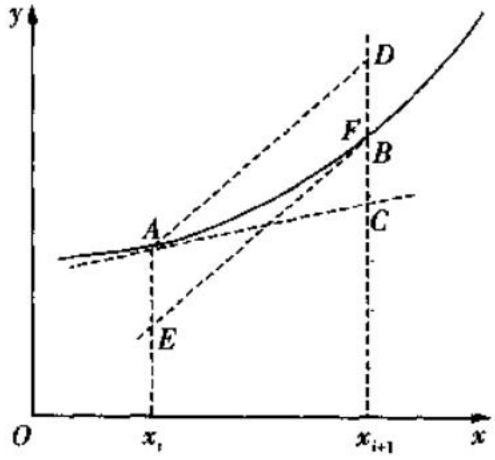


图8-2

$$\begin{cases} y_{i+1}=y_i+\dfrac{1}{2}h(k_1+k_2) \\ k_1=f(x_i,y_i) \\ k_2=f(x_i+h,y_i+k_1h) \end{cases} \qquad (i=0,1,2,\cdots,n-1)$$

例 8-1 已知在管式反应器中有液相吸热反应 $A \longrightarrow R+S$，反应管外油浴温度为 340 ℃，假定管内温度与转化率的关系为

$$\frac{\mathrm{d}t}{\mathrm{d}x_A}=-65.0-\frac{15.58(t-t_c)}{k(1-x_A)}$$

其中，速率常数 $k=1.17\times10^{17}\exp\left(-\dfrac{22\,143.94}{T}\right)$，$t_c$ 为反应器外壁温度。

若反应器入口温度为 340 ℃，反应器出口转化率为 90%，试用改进尤拉公式求反应器出口温度。

解 根据题意可知此题为初值问题

$$\begin{cases} \dfrac{\mathrm{d}t}{\mathrm{d}x_A}=-65.0-\dfrac{15.58(t-t_c)}{k(1-x_A)} \\ t\,|_{x_A=0}=340\ ℃ \end{cases}$$

假定反应器外壁温度恒定，且等于油浴温度，则

$$\frac{\mathrm{d}t}{\mathrm{d}x_A}=-65.0-\frac{15.58(t-340)}{k(1-x_A)}$$

$$k=1.17\times10^{17}\exp\left(-\frac{22\,143.94}{t+273.15}\right)$$

取步长 $h=\Delta x_A=0.05$，用改进尤拉公式

$$\begin{cases} t_{i+1}^{(0)}=t_i+hf(x_{Ai},t_i) \\ t_{i+1}^{(k+1)}=t_i+\dfrac{h}{2}[f(x_{Ai},t_i)+f(x_{Ai+1},t_{i+1}^{(k)})] \\ x_{Ai+1}=x_{A0}+(i+1)h \end{cases} \quad (k=0,1,2,\cdots;i=0,1,2,\cdots,n-1)$$

从 $x_{A0}=0$，$t_0=340$ ℃开始计算，计算结果如下：

x_A	0	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
t	340	336.8	333.91	331.32	329.10	327.02	326.02	325.10	325.00	324.81
x_A	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	
t	325.13	325.58	326.16	326.90	327.61	328.56	329.83	331.30	333.01	

四、尤拉两步公式

若应用中心差商近似代替任意点 x_i 处的导数，即

$$\frac{dy}{dx}_{x_i} \approx \frac{y(x_{i+1}) - y(x_{i-1})}{2h}$$

则

$$y(x_{i+1}) \approx y(x_{i-1}) + 2hf(x_i, y_i)$$

若令 $y_{i-1}, y_i \approx y(x_i), y_{i+1} \approx y(x_{i+1})$，则

$$\begin{cases} y_{i+1} = y_{i-1} + 2hf(x_i, y_i) \\ x_{i+1} = x_0 + (i+1)h \end{cases}$$

其中，$y_1 = y_0 + hf(x_0, y_0)$；$i = 1, 2, \cdots, n-1$。

此式称为尤拉两步公式，也称作中心点法，是一种显式公式，即从(x_0, y_0)出发，可由显式尤拉法提供 y_1，再由 y_{i-1} 和 y_i 计算出 y_{i+1}。可以证明其局部截断误差为 $o(h^3)$，总误差为 $o(h^2)$。计算精度比显式尤拉公式高一阶。但它的精度的提高是由多提供信息换来的。

例 8-2 分别应用梯形公式和尤拉两步公式求解下列初值问题

$$\begin{cases} \dfrac{dy}{dx} = -y + x^2 + 1, \quad x \in [0,1] \\ y(0) = 1 \end{cases}$$

解 取步长 $h = 0.1$，则其计算结果如下：

x_i	0	0.1	0.2	0.3	0.4	0.5
梯形公式 y_i	1.000 000	1.000 500	1.002 912	1.008 926	1.020 128	1.037 916
两步公式 y_i	1.000 000	1.000 250	1.002 426	1.008 245	1.019 262	1.036 882
精确解 y_i	1.000 000	1.000 325	1.002 538	1.008 363	1.019 359	1.036 938

x_i	0.6	0.7	0.8	0.9	1.0
梯形公式 y_i	1.063 564	1.098 225	1.142 944	1.198 664	1.266 241
两步公式 y_i	1.062 378	1.096 902	1.141 496	1.197 104	1.264 579
精确解 y_i	1.062 376	1.096 829	1.141 342	1.196 860	1.264 241

可以看出，尤拉两步公式的准确度比梯形公式还好，但它不太稳定，可以证明，当 i 很大时，舍入误差积累很快，会影响计算值精度。

第二节 解初值问题的龙格—库塔法

由改进尤拉公式可知，增加函数$f(x,y)$的计算次数，可提高方法的精度。这样，若再增加函数的计算次数，方法的精度将还会提高，龙格—库塔法就是基于这种思想推导出

来的。

梯形公式精度高的根本原因是利用了 x_i 处的斜率 $f(x_i,y_i)$ 及 x_{i+1} 处的斜率 $f(x_{i+1},y_{i+1})$ 的平均值，这样，在区间 $[x_i,x_{i+1}]$ 内多取几个点处的斜率，然后以它们的加权平均值作为区间 $[x_i,x_{i+1}]$ 上的平均值，构造出精度更高的计算公式。这便是龙格—库塔法的基本思想。

龙格—库塔法也是一种显式算法，它的一般形式为

$$y_{i+1}=y_i+h\sum_{j=1}^{m}\lambda_j K_j,$$

$$K_j=f\left(x_i+\alpha_j h,y_i+h\sum_{k=1}^{j-1}\beta_{jk}K_k\right)$$

$$(i=0,1,2,\cdots,n-1;\quad j=1,2,\cdots,m)$$

其中，m 为计算次数，λ_j 为权系数，α_j 为龙格—库塔法节点，$\alpha_1=0$。

若 $m=1$，$\lambda=1$，则 $k_1=f(x_i,y_i)$，于是就有

$$y_{i+1}=y_i+hf(x_i,y_i)$$

这便是显式尤拉公式，所以显式尤拉公式是最低阶的龙格—库塔公式。

当 $m=2$，即为二阶龙格—库塔公式，则

$$\begin{cases}y_{i+1}=y_i+h(\lambda_1K_1+\lambda_2K_2)\\K_1=f(x_i,y_i)\\K_2=f(x_i+\alpha_2h,y_i+\beta_{21}K_1)\end{cases}$$

要得到具体的表达式，就应求出 $\lambda_1,\lambda_2,\alpha_2,\beta_{21}$。

将 $y(x)$ 在 x_i 处作泰勒展开有

$$y(x_{i+1})=y(x_i)+hf(x_i,y(x_i))+\frac{h^2}{2}f'(x_i,y(x_i))+o(h^3)$$

而

$$f'(x_i,y(x_i))=\left[\frac{\partial f}{\partial x}+\frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}x}\right]_{x=x_i}=(f_x+f_yf)_{x=x_i}$$

这样，略去三阶小量 $o(h^3)$，取 $y_{i+1}\approx y(x_{i+1})$，$y_i\approx y(x_i)$，则

$$\begin{aligned}y_{i+1}&=y_i+hf(x_i,y(x_i))+\frac{h^2}{2}(f_x+f_yf)_{x=x_i}\\&=y_i+hf(x_i,y_i)+\frac{h^2}{2}f_x(x_i,y_i)+\frac{h^2}{2}f_y(x_i,y_i)f(x_i,y_i)\end{aligned}$$

又

$$K_1=f(x_i,y_i),$$

$$\begin{aligned}K_2&=f(x_i+\alpha_2h,y_i+\beta_{21}K_1)\\&=f(x_i,y_i)+\alpha_2hf_x(x_i,y_i)+\beta_{21}f(x_i,y_i)f_y(x_i,y_i)\end{aligned}$$

于是二阶龙格—库塔公式可写为

$$y_{i+1}=y_i+(\lambda_1+\lambda_2)hf(x_i,y_i)+\lambda_2\alpha_2h^2f_x(x_i,y_i)+\lambda_2\beta_{21}h^2f(x_i,y_i)f_y(x_i,y_i)$$

从而

$$hf(x_i,y_i)+\frac{1}{2}h^2f_x(x_i,y_i)+\frac{1}{2}h^2f_y(x_i,y_i)f(x_i,y_i)$$

$$=(\lambda_1+\lambda_2)hf(x_i,y_i)+\lambda_2\alpha_2h^2f_x(x_i,y_i)+\lambda_2\beta_{21}h^2f_y(x_i,y_i)f(x_i,y_i)$$

比较等式两边可得

$$\begin{cases}\lambda_1+\lambda_2=1\\ \lambda_2\alpha_2=\dfrac{1}{2}\\ \lambda_2\beta_{21}=\dfrac{1}{2}\end{cases}$$

这是一个有4个未知数和3个方程组成的方程组，它有无穷多解。凡满足这一条件的一族公式均称为二阶龙格—库塔公式。

显然，当 $\lambda_1=\lambda_2=\frac{1}{2}$，$\alpha_2=\beta_{21}=1$ 时

$$y_{i+1}=y_i+\frac{1}{2}[f(x_i,y_i)+f(x_{i+h},y_i+hK_1)]$$

此即为改进尤拉公式。

同理，可以推出同一族三阶的龙格—库塔公式，其中最常用的形式为

$$\begin{cases}y_{i+1}=y_i+\dfrac{1}{6}h(K_1+4K_2+K_3)\\ K_1=f(x_i,y_i)\\ K_2=f\left(x_i+\dfrac{1}{2}h,y_i+\dfrac{1}{2}hK_1\right)\\ K_3=f(x_i+h,y-hK_1+2hK_2)\end{cases}$$

其局部截断误差为 $o(h^4)$，总误差为 $o(h^3)$。

一般来说，随着计算次数 m 的增大，计算的精度阶增高。但是，推导证明，方法的精度阶不总是与函数计算次数成比例。龙格—库塔法的精度阶请参阅表8－1。

表8－1

函数计算次数	1	2	3	4	5	6	7	≥8
局部截断误差	$o(h^2)$	$o(h^3)$	$o(h^4)$	$o(h^5)$	$o(h^5)$	$o(h^6)$	$o(h^7)$	$o(h^{m+1})$
方法的精度阶	1	2	3	4	4	5	6	$m-2$

由表8－1可以看出，取 $m=4$ 是最合适的。因此，目前最常用的是四阶龙格—库塔法，也称为经典龙格—库塔法。其形式为

$$\begin{cases} y_{i+1}=y_i+\dfrac{h}{6}(K_1+2K_2+2K_3+K_4) \\ K_1=f(x_i,y_i) \\ K_2=f\left(x_i+\dfrac{1}{2}h,y_i+\dfrac{1}{2}hK_1\right) \\ K_3=f\left(x_i+\dfrac{1}{2}h,y_i+\dfrac{1}{2}hK_2\right) \\ K_4=f(x_i+h,y_i+hK_3) \end{cases}$$

其中，$i=0,1,2,\cdots,n-1$。

与一般式相比较，经典龙格—库塔法的 $m=4$，此时 $\lambda_1=\lambda_4=\dfrac{1}{6}$，$\lambda_2=\lambda_3=\dfrac{1}{3}$，$\alpha_2=\alpha_3=\beta_{21}=\beta_{32}=\dfrac{1}{2}$，$\alpha_4=\beta_{43}=1$，$\beta_{31}=\beta_{41}=\beta_{42}=0$。由表 8-1 可知，该法计算函数 4 次，精度阶为 4 阶。显然，在 $[x_i,x_{i+1}]$ 内取 3 个节点 x_i，$x_{i+\frac{1}{2}}$，x_{i+1}，而在 $x_{i+\frac{1}{2}}$ 处计算 2 次函数值。因此，不增设节点，只增加函数计算次数，也可达到提高精度的目的。

经典龙格—库塔法不仅取最合适的 $m=4$，而且计算过程规律性强，便于计算机计算。为编程方便，令 $e_1=e_2=e_5=\dfrac{1}{2}h$，$e_3=e_4=h$，则有

$$\begin{cases} y_{i+1}=y_i+\sum\limits_{j=1}^{4}e_{j+1}K_j/3 \\ K_j=f(x,y) \\ x=x_i+e_j \\ y=y_i+e_jK_j \end{cases}$$

其中 $i=0,1,2,\cdots,n-1$；$j=1,2,3,4$。

最后应指出的是，在实际计算中选取合适的步长 h 是很重要的，h 小时有利于降低局部截断误差，提高计算精度；但 h 过小，不仅使计算步骤增多，计算工作量加大，还会导致舍入误差的大量积累，反而使后部的计算精度降低。

由于事先估计步长 h 较为困难，故在实际计算中采用变步长的方法。对于经典龙格—库塔法，由于它是四阶，从 x_i 出发，以 h 为步长，经过一步计算求得 $y(x_{i+1})$ 的近似值 $y_{i+1}^{(h)}$，其局部截断误差应为

$$y(x_{i+1})-y_{i+1}^{(h)}=Ch^5$$

其中，C 为系数，当 h 不太大时，可视为常数。

然后将步长减半，即取 $\dfrac{h}{2}$，再由 x_i 出发，经两步计算求得 $y(x_{i+1})$ 的近似值 $y_{i+1}^{(\frac{h}{2})}$，其每步计算的局部截断误差应为 $C\left(\dfrac{h}{2}\right)^5$，则

$$y(x_{i+1}) - y_{i+1}^{\left(\frac{h}{2}\right)} = 2C\left(\frac{h}{2}\right)^5$$

于是

$$\frac{y(x_{i+1}) - y_{i+1}^{\left(\frac{h}{2}\right)}}{y(x_{i+1}) - y_{i+1}^{(h)}} = \frac{1}{16}$$

则

$$y(x_{i+1}) - y_{i+1}^{\left(\frac{h}{2}\right)} = \frac{1}{15}\left(y_{i+1}^{\left(\frac{h}{2}\right)} - y_{i+1}^{(h)}\right)$$

这表明，以 $y_{i+1}^{\left(\frac{h}{2}\right)}$ 作为 $y(x_{i+1})$ 的近似值，其计算误差可以前后两次变步长计算的结果之差来表示。因此，在变步长计算中只需要考察

$$\left|y_{i+1}^{\left(\frac{h}{2}\right)} - y_{i+1}^{(h)}\right| \leqslant \varepsilon$$

是否成立，若成立，则取 $y_{i+1}^{\left(\frac{h}{2}\right)}$ 为 $y(x_{i+1})$ 的近似值；否则，将步长再次减半进行计算，直到满足精度要求，以 $y_{i+1}^{\left(\frac{h}{2}\right)}$ 为计算结果。这种变步长的算法称之为步长的自动选择，这在计算机上是不难实现的。

例 8-3 应用经典龙格—库塔法计算初值问题：

$$\begin{cases} \dfrac{dy}{dx} = y - \dfrac{2x}{y}, & x \in [0,1] \\ y(0) = 1 \end{cases}$$

解 此题的解析解为 $y = \sqrt{1+2x}$。现取步长 $h = 0.1$，将由经典龙格—库塔法和精确解的计算结果列表如下：

x_i	0	0.1	0.2	0.3	0.4	0.5
龙格—库塔法 y_i	1	1.095 446	1.183 217	1.264 912	1.341 642	1.141 414
精确解 y_i	1	1.095 445	1.183 216	1.264 911	1.341 641	1.414 214

x_i	0.6	0.7	0.8	0.9	1.0
龙格—库塔法 y_i	1.483 242	1.549 196	1.612 455	1.673 325	1.732 056
精确解 y_i	1.483 240	1.549 193	1.612 452	1.673 320	1.732 051

可见经典龙格—库塔法的精度较高，同时随 x_i 的增大，误差 $|y_i - y(x_i)|$ 也增大，这说明其截断误差是逐步累积的。

第三节 解初值问题的线性多步法

显式尤拉法和龙格—库塔法均为单步法，单步法是指只要给出初始 y_0，就可顺序地

算出 $y_1, y_2, \cdots, y_n$，即其是自行起步的，而多步法则除了已知的 y_0 外，还要借助其他单步法提供 $y_1, y_2, \cdots, y_i$ 等若干个节点值，即它不是自行起步的。

由尤拉两步公式的讨论可知，增加信息量也可提高方法的精度。多步法便是受此思想的启发而提出的，即再通过增加信息量，使方法的精度更高。例如，在计算 y_{i+1} 时，充分利用已知 $y_1, y_2, \cdots, y_i$ 以及 $f(x_0, y_0), f(x_1, y_1), \cdots f(x_i, y_i)$，从而使 y_{i+1} 的计算精度提高。由于保留并利用了前面几步所取得的信息，因此，该法称为多步法。多步法的通式为

$$y_{i+1} = \sum_{j=0}^{m-1} \alpha_j y_{i-j} + h \sum_{j=-1}^{m-1} b_j f(x_{i-j}, y_{i-j}) \quad (i = m-1, m, \cdots, n)$$

其中，m 为大于 1 的正整数；$y_1, y_2, \cdots, y_{m-1}$ 由单步法提供或者已知。

阿达姆斯法是多步法中最常用的一种形式，其通式为

$$y_{i+1} = y_i + h \sum_{j=-1}^{m-1} b_j^{(m)} f(x_{i-j}, y_{i-j}) \quad (i = m-1, m, m+1, \cdots, n)$$

其中，m 为大于 1 的正整数，表示利用了前 m 步的已知信息。此式称为 m 步阿达姆斯公式。当 $b_{-1}^{(m)} \neq 0$ 时，公式中含有 $f(x_{i+1}, y_{i+1})$，为隐式形式，需迭代求解；当 $b_{-1}^{(m)} = 0$ 时，公式中不含有 $f(x_{i+1}, y_{i+1})$，为显式公式，可直接解出 y_{i+1}。

多步法的基本思路是将常微分方程的初值问题转化为等价的积分问题，然后，应用插值多项式代替被积函数 $f(x, y)$，从而使微分方程离散化。

设有初值问题

$$\begin{cases} \dfrac{dy}{dx} = f(x, y), & x \in [x_0, x_n] \\ y|_{x=x_0} = y_0 \end{cases}$$

其等价的积分方程为

$$y(x) = y_0 + \int_{x_0}^{x_n} f(x, y)\, dx$$

取定节点 $x_0, x_1, \cdots, x_n$，则有

$$\begin{cases} y(x_{i+1}) = y(x_i) + \displaystyle\int_{x_i}^{x_{i+1}} f(x, y)\, dx \quad (i = 0, 1, 2, \cdots, n-1) \\ y(x_0) = y_0 \end{cases}$$

因积分通常不易直接积出，需要用插值多项式 $P_m(x)$ 代替 $f(x, y)$，从而有

$$y(x_{i+1}) = y(x_i) + \int_{x_i}^{x_{i+1}} P_m(x)\, dx$$

应用不同的数值积分法或不同的项数 m，可获得不同的多步法公式。这里主要讨论采用插值型求积方法的阿达姆斯公式。

一、阿达姆斯隐式公式

以阿达姆斯三步隐式公式为例进行推导，即取 $m=3$。若利用已知点 (x_{i-2},y_{i-2})，(x_{i-1},y_{i-1})，(x_i,y_i) 及待求点 (x_{i+1},y_{i+1}) 构造插值多项式 $P_3(x)$，则有

$$P_3(x)=\sum_{k=i-2}^{i+1}\left[\prod_{\substack{j=i-2\\ j\neq k}}^{i+1}\frac{x-x_k}{x_k-x_j}f(x_k,y_k)\right]$$

于是

$$y(x_{i+1})\approx y(x_i)+\sum_{k=i-2}^{i+1}\int_{x_i}^{x_{i+1}}\left[\prod_{\substack{j=i-2\\ j\neq k}}^{i+1}\frac{x-x_k}{x_k-x_j}f(x_k,y_k)\right]\mathrm{d}x$$

取

$$y(x_{i+1})\approx y_{i+1},\quad y(x_i)\approx y_i,\quad y(x_{i-1})\approx y_{i-1},\quad y(x_{i-2})\approx y_{i-2}$$

由于节点为等距分布 $h=\dfrac{x_n-x_0}{n}$，$x_i=x_0+ih\ (i=0,1,\cdots,n)$

令

$$x=x_i+ht,\quad \mathrm{d}x=h\mathrm{d}t$$

当 $x=x_i$，时，$t=0$；当 $x=x_{i+1}$ 时，$t=1$，则

$$y_{i+1}=y_i+\frac{h}{6}f(x_{i+1},y_{i+1})\int_0^1 t(t+1)(t+2)\mathrm{d}t-\frac{h}{2}f(x_i,y_i)\int_0^1(t-1)(t+1)(t+2)\mathrm{d}t+$$

$$\frac{h}{2}f(x_{i-1},y_{i-1})\int_0^1(t-1)t(t+2)\mathrm{d}t-\frac{h}{6}f(x_{i-2},y_{i-2})\int_0^1(t-1)t(t+1)\mathrm{d}t$$

从而可得到

$$y_{i+1}=y_i+\frac{h}{24}[9f(x_{i+1},y_{i+1})+19f(x_i,y_i)-5f(x_{i-1},y_{i-1})+f(x_{i-2},y_{i-2})]$$

$$(i=2,3,\cdots,n)$$

此式称为阿达姆斯三步隐式公式。它利用了前 3 步的已知信息 y_{i-2},y_{i-1},y_i。由于它是隐式公式，需迭代求解。

现讨论阿达姆斯三步隐式公式的局部截断误差。

由于

$$R_3(x)=\frac{1}{4!}f^{(4)}(\xi_i)\prod_{k=i-2}^{i+1}(x-x_k)\quad (x_i\leqslant\xi_i\leqslant x_{i+1})$$

则

$$\begin{aligned} y(x_{i+1}) - y_{i+1} &= \frac{1}{4!} f^{(4)}(\xi_i) \int_{x_i}^{x_{i+1}} \prod_{k=i-2}^{i+1} (x - x_k)\,\mathrm{d}x \\ &= -\frac{19}{720} h^5 f^{(4)}(\xi_i) \\ &= o(h^5) \end{aligned}$$

其局部截断误差为 $o(h^5)$，总误差为 $o(h^4)$，具有四阶精度。

利用类似的方法可导出阿达姆斯 m 步隐式公式。现将不同的 $m(m=0,1,2,3,4,5)$ 值下的系数 $b_j^{(m)}(j=-1,0,1,2,\cdots,m-1)$ 列于表 8-2。

表 8-2

误差阶	$(m)/j$	-1	0	1	2	3	4
$o(h^2)$	0 $b_j^{(0)}$	1					
$o(h^3)$	1 $b_j^{(1)}$	$\frac{1}{2}$	$\frac{1}{2}$				
$o(h^4)$	2 $b_j^{(2)}$	$\frac{5}{12}$	$\frac{8}{12}$	$-\frac{1}{12}$			
$o(h^5)$	3 $b_j^{(3)}$	$\frac{9}{24}$	$\frac{19}{24}$	$-\frac{5}{24}$	$\frac{1}{24}$		
$o(h^6)$	4 $b_j^{(4)}$	$\frac{251}{720}$	$\frac{646}{720}$	$-\frac{264}{720}$	$\frac{106}{720}$	$-\frac{19}{720}$	
$o(h^7)$	5 $b_j^{(5)}$	$\frac{475}{1\,440}$	$\frac{1\,427}{1\,440}$	$-\frac{798}{1\,440}$	$\frac{482}{1\,440}$	$-\frac{173}{1\,440}$	$\frac{27}{1\,440}$

由表 8-2 可见，步数 m 越多，公式的局部截断误差越小，精度越高。当然，计算机内存的占用也越多。

还应指出的是，尽管阿达姆斯法与尤拉法的推导思路不尽相同，但两者却有着内在的联系。当 $m=0$ 时，阿达姆斯隐式公式便为隐式尤拉公式，即

$$y_{i+1} = y_i + hf(x_{i+1}, y_{i+1})$$

当 $m=1$ 时，阿达姆斯隐式公式便为梯形公式，即

$$y_{i+1}=y_i+\frac{1}{2}h[f(x_i,y_i)+f(x_{i+1},y_{i+1})]$$

二、阿达姆斯显式公式

若利用已知点(x_{i-3},y_{i-3}),(x_{i-2},y_{i-2}),(x_{i-1},y_{i-1})和(x_i,y_i),则可推导出阿达姆斯四步显式公式,其精度可达四阶。

由已知的4个信息点,可构造三次多项式

$$P_3(x)=\sum_{k=i-3}^{i}\left[\prod_{\substack{j=i-3\\ j\neq k}}^{i}\frac{x-x_k}{x_k-x_j}f(x_k,y_k)\right]$$

于是

$$y_{i+1}=y_i+\sum_{k=i-3}^{i}\int_{x_i}^{x_{i+1}}\left[\prod_{\substack{j=i-3\\ j\neq k}}^{i}\frac{x-x_k}{x_k-x_j}f(x_k,y_k)\right]\mathrm{d}x$$

令

$$x=x_i+ht,\quad \mathrm{d}x=h\mathrm{d}t$$

当$x=x_i$时,$t=0$,当$x=x_{i+1}$时,$t=1$,则

$$y_{i+1}=y_i-\frac{h}{6}f(x_{i-3},y_{i-3})\int_0^1 t(t+1)(t+2)\mathrm{d}t+\frac{1}{2}f(x_{i-2},y_{i-2})\int_0^1 t(t+1)(t+3)\mathrm{d}t-$$

$$\frac{h}{2}f(x_{i-1},y_{i-1})\int_0^1 t(t+2)(t+3)\mathrm{d}t+\frac{h}{6}f(x_i,y_i)\int_0^1 t(t+1)(t+2)(t+3)\mathrm{d}t$$

于是,积分后得

$$y_{i+1}=y_i+\frac{h}{24}[55f(x_i,y_i)-59f(x_{i-1},y_{i-1})+37f(x_{i-2},y_{i-2})-9f(x_{i-3},y_{i-3})]$$

$$(i=3,4,\cdots,n-1)$$

该式称为阿达姆斯四步显式公式。由于$b_{-1}^{(4)}=0$公式中不含有$f(x_{i+1},y_{i+1})$项,为外推公式,可直接解出y_{i+1}。

同样,由$R_3(x)=\frac{1}{4!}f^{(4)}(\xi_i)\prod_{j=i-3}^{i}(x-x_j)(x_i\leqslant\xi_i\leqslant x_{i+1})$可得阿达姆斯四步显式公式的局部截断误差

$$y(x_{i+1})-y_{i+1}=\frac{1}{4!}f^{(4)}(\xi_i)\int_{x_i}^{x_{i+1}}\prod_{j=i-3}^{i}(x-x_j)\mathrm{d}x$$

$$=\frac{251}{720}h^5f^{(4)}(\xi_i)=o(h^5)$$

因而其总误差为 $o(h^4)$，具有四阶精度。

类似地，可导出其他 m 值的阿达姆显式公式，表 8－3 列出了部分公式的系数 $b_j^{(m)}$ $(j=0,1,2,\cdots,m-1)$。

表 8－3

误差阶	$(m)/j$	0	1	2	3	4	5
$o(h^2)$	1 $b_j^{(1)}$	1					
$o(h^3)$	2 $b_j^{(2)}$	$\frac{3}{2}$	$-\frac{1}{2}$				
$o(h^4)$	3 $b_j^{(3)}$	$\frac{23}{12}$	$-\frac{16}{12}$	$\frac{5}{12}$			
$o(h^5)$	4 $b_j^{(4)}$	$\frac{55}{24}$	$-\frac{59}{24}$	$\frac{37}{24}$	$-\frac{9}{24}$		
$o(h^6)$	5 $b_j^{(5)}$	$\frac{1\,901}{720}$	$-\frac{2\,774}{720}$	$\frac{2\,616}{720}$	$-\frac{1\,247}{720}$	$\frac{251}{720}$	
$o(h^7)$	6 $b_j^{(6)}$	$\frac{4\,277}{1\,440}$	$-\frac{7\,923}{1\,440}$	$\frac{9\,982}{1\,440}$	$-\frac{7\,298}{1\,440}$	$\frac{2\,877}{1\,440}$	$-\frac{475}{1\,440}$

显然，各显式公式均为外推公式，可直接解出 y_{i+1}，且 m 越大，精度越高。当 $m=1$ 时，即为显式尤拉公式 $y_{i+1}=y_i+hf(x_i,y_i)$。

现将阿达姆斯显式公式与隐公式相比较可得出：

(1)在相同的误差阶下，显式公式要比隐式公式多使用一个已知信息。例如，在 $o(h^5)$ 下，显式公式利用了已知的 y_{i-3},y_{i-2},y_{i-1} 和 y_i，而隐式公式则只用 y_{i-2},y_{i-1} 和 y_i，并且显式公式的局部截断误差的主部系数比隐式公式的大，计算误差大。如 $o(h^5)$ 下，显式公式局部截断误差的主部系数为 $\frac{251}{720}$，而隐式公式的仅为 $-\frac{19}{720}$。

(2)在相同误差阶下，显式公式的计算工作量比隐式公式的少。显式公式可直接求解，而隐式公式要迭代求解。

因此，在实际计算时，通常是将两者结合使用，构成预报—校正格式。

三、阿达姆斯预报—校正格式

阿达姆斯预报—校正格式就是由显式公式提供初值，再由同阶的隐式公式进行校正。这既保持了隐式公式精度高的优点，又能为它提供较好的初值，减少迭代次数。在一般情况下，迭代 1 次就可使精度得到明显改进，因此，在实际计算中大多采用预报 1

次，校正1次的格式。四阶阿达姆斯预报—校正格式如下：

$$\begin{cases} y_{i+1}^{0} = y_i + \dfrac{h}{24}[55f(x_i,y_i) - 59f(x_{i-1},y_{i-1}) + 37f(x_{i-2},y_{i-2}) - 9f(x_{i-3},y_{i-3})] \\ y_{i+1} = y_i + \dfrac{h}{24}[9f(x_{i+1},y_{i+1}^{0}) + 19f(x_i,y_i) - 5f(x_{i-1},y_{i-1}) + f(x_{i-2},y_{i-2})] \end{cases}$$

其中，$i=3,4,\cdots,n-1$。

显然，每跨一步，只需计算两次函数值，即计算$f(x_i,y_i)$和$f(x_{i+1},y_{i+1}^{0})$，要比同阶的龙格—库塔法计算工作量少。但它在计算过程中要保留较多的已知信息，即多使用内存。例如，对于四阶阿达姆斯法来说，计算y_{i+1}时，必须保留在计算y_i时已知的3个信息$f(x_{i-1},y_{i-1})$、$f(x_{i-2},y_{i-2})$、$f(x_{i-3},y_{i-3})$，再者，阿达姆斯法是多步法，它不能自起步，必须由单步法提供开头几个函数值，对四阶阿达姆斯法，一般由四阶龙格—库塔法提供y_1、y_2、y_3。

预报—校正格式的误差估计较为方便，对于四阶显式和隐式公式的局部截断误差分别为

$$y(x_{i+1}) = y_{i+1}^{0} + \frac{251}{720}h^5 f^{(4)}(\xi_1)$$

$$y(x_{i+1}) = y_{i+1} - \frac{19}{720}h^5 f^{(4)}(\xi_2)$$

当h较小，$f^{(4)}(x,y)$在$[x_i,y_i]$上变化不大时，近似地取$f^{(4)}(\xi_1)\approx f^{(4)}(\xi_2)$，则

$$y_{i+1} - y_{i+1}^{0} = \frac{270}{720}h^5 f^{(4)}(\xi_1)$$

于是

$$y(x_{i+1}) - y_{i+1} = -\frac{19}{270}(y_{i+1} - y_{i+1}^{0})$$

则

$$|y(x_{i+1}) - y_{i+1}| = \frac{19}{270}|y_{i+1} - y_{i+1}^{0}|$$

这样，可用校正值与预报值之差$|y_{i+1} - y_{i+1}^{0}|$来估计近似值y_{i+1}的误差$|y(x_{i+1}) - y_{i+1}|$。

四阶阿达姆斯预报—校正格式的计算步骤为：

(1)给定区间端点x_0，x_n，初值y_0，步长h和要求精度ε。

(2)由四阶龙格—库塔法计算y_1，y_2，y_3，并存留$f(x_1,y_1)$、$f(x_2,y_2)$、$f(x_3,y_3)$。

(3)对于$i=3,4,\cdots,n-1$，计算

$$x_i = x_0 + ih,$$

$$y_{i+1}^{0} = y_i + \frac{h}{24}[55f(x_i,y_i) - 59f(x_{i-1},y_{i-1}) + 37f(x_{i-2},y_{i-2}) - 9f(x_{i-3},y_{i-3})],$$

$$y_{i+1} = y_i + \frac{h}{24}[9f(x_{i+1},y_{i+1}^{0}) + 19f(x_i,y_i) - 5f(x_{i-1},y_{i-1}) + f(x_{i-2},y_{i-2})]$$

(4)若$|y_{i+1}-y_{i+1}^0|\leqslant\varepsilon$,则输出结果,终止计算;否则,令$h=h/2$,回第(2)步。

例 8-4 已知气流通过固定床的压降公式为

$$\frac{dp}{dl}=-\frac{10.63}{p}$$

若床层高度$l=3$ m,床层入口压力$p_0=15$ atm试用四阶阿达姆斯预报—校正格式求床层内的压力分布。

解 由题意可知,初值问题为

$$\begin{cases}\dfrac{dp}{dl}=-\dfrac{10.63}{p} & l\in[0,3]\\ l=0, & p(0)=15\end{cases}$$

选用步长$h=\Delta l=0.5$,用四阶龙格—库塔法求得初始值为$p_1=14.64$,$p_2=14.27$,$p_3=13.90$。然后由四阶阿达姆斯预报—校正格式

$$\begin{cases}p_{i+1}^0=p_i-\dfrac{10.63}{24}h\left[\dfrac{55}{p_i}-\dfrac{59}{p_{i-1}}+\dfrac{37}{p_{i-2}}-\dfrac{9}{p_{i-3}}\right]\\ p_{i+1}=p_i-\dfrac{10.63}{24}h\left[\dfrac{9}{p_{i+1}^0}+\dfrac{19}{p_i}-\dfrac{5}{p_{i-1}}+\dfrac{1}{p_{i-2}}\right]\end{cases}$$

进行计算,计算结果为:

l/m	0	0.5	1.0	1.5	2.0	2.5	3.0
p/atm	15	14.64	14.27	13.90	13.51	13.11	12.70

第四节 常微分方程组初值问题的数值解法

一阶常微分方程组的初值问题可表示为

$$\begin{cases}\dfrac{dy_1}{dx}=f_1(x,y_1,y_2\cdots,y_m)\\ \dfrac{dy_2}{dx}=f_2(x,y_1,y_2\cdots,y_m)\\ \cdots\cdots\\ \dfrac{dy_m}{dx}=f_m(x,y_1,y_2\cdots,y_m)\\ y_1(x_0)=y_{10}\\ y_2(x_0)=y_{20}\\ \cdots\cdots\\ y_m(x_0)=y_{m0}\end{cases}$$

其中，$x \in [x_0, x_n]$。

即

$$\begin{cases} \dfrac{dy_j}{dx} = f_j(x, y_1, y_2 \cdots, y_m), & x \in [x_0, x_n] \\ y_j(x_0) = y_{j0}, & j = 1, 2, \cdots, m \end{cases}$$

写为向量形式即为

$$\begin{cases} \dfrac{dY}{dx} = F(x, Y), & x \in [x_0, x_n] \\ Y(x_0) = Y_0 \end{cases}$$

其中，$Y = (y_1, y_2, \cdots, y_m)^T$，$Y_0 = (y_{10}, y_{20}, \cdots, y_{m0})^T$，$F = (f_1, f_2, \cdots f_m)^T$。

显然，常微分方程组的初值问题是微分方程初值问题的扩展。因此，可用任何解常微分方程初值问题的方法来解常微分方程组的初值问题。例如，在实际计算中常用的龙格—库塔法和阿达姆斯法。

一、龙格—库塔法

经典龙格—库塔法用于求解常微分方程组的初值问题时可表达为

$$\begin{cases} y_{j,i+1} = y_{ji} + \dfrac{1}{6}h(K_{j1} + 2K_{j2} + 2K_{j3} + K_{j4}) \\ K_{j1} = f_j(x_i, y_{1i}, y_{2i}, \cdots, y_{mi}) \\ K_{j2} = f_j(x_i + \dfrac{h}{2}, y_{1i} + \dfrac{h}{2}K_{11}, y_{2i} + \dfrac{h}{2}K_{21}, \cdots, y_{mi} + \dfrac{h}{2}K_{m1}) \\ K_{j3} = f_j(x_i + \dfrac{h}{2}, y_{1i} + \dfrac{h}{2}K_{12}, y_{2i} + \dfrac{h}{2}K_{22}, \cdots, y_{mi} + \dfrac{h}{2}K_{m2}) \\ K_{j4} = f_j(x_i + h, y_{1i} + hK_{13}, y_{2i} + hK_{23}, \cdots, y_{mi} + hK_{m3}) \\ j = 1, 2, \cdots, m; i = 0, 1, 2, \cdots, n-1 \end{cases}$$

其中，$h = \dfrac{x_n - x_0}{n}$；$x_i = x_0 + ih$ $(i = 0, 1, 2, \cdots, n)$。

为了编程方便，通常令 $y_0 = x$，那么有

$$\frac{dy_0}{dy_0} = \frac{dy_0}{dx} = f_0(y_0) \equiv 1$$

使原 m 个方程联立的方程组变为 $m+1$ 个方程联立的方程组

$$\begin{cases} \dfrac{dy_j}{dy_0} = f_j(y_0, y_1, y_2, \cdots, y_m), & y_0 \in [x_0, x_n] \\ y_j(y_0) = y_{j0}, & j = 0, 1, 2, \cdots, m \end{cases}$$

其中，$f_0(y_0, y_1, y_2, \cdots, y_m) = 1$，$y_0(y_0) = y_{00} = x_0$，实质上第零个方程的解 $y_0(x)$ 便是 x_0。

对该式应用经典龙格—库塔法，并取 $e_1 = e_2 = e_5 = \dfrac{1}{2}h$，$e_3 = e_4 = h$，则有

$$\begin{cases} y_{ji+1} = y_{ji} + \sum_{k=1}^{4} e_{k+1} K_{jk}/3 \\ K_{jk} = f_j(y_0, y_1, y_2, \cdots, y_m) \\ y_j = y_{ji} + e_k K_{jk} \end{cases}$$

其中 $j=0,1,2,\cdots,m$；$i=0,1,2,\cdots,n-1$；$k=1,2,3,4$。

例 8-5 萘在列管式固定床反应器中生产苯酐。已知萘的转化率和反应温度沿床层的变化率为

$$\frac{\mathrm{d}x}{\mathrm{d}l} = 1.1094 \times 10^{11}\left(\frac{1-x}{1\,000-0.5x}\right)^{0.38} \exp\left(-\frac{14\,098}{T}\right)$$

$$\frac{\mathrm{d}T}{\mathrm{d}l} = 7.6120 \times 10^{12}\left(\frac{1-x}{1\,000-0.5x}\right)^{0.38} \exp\left(-\frac{14\,098}{T}\right) - 9.9282(T-T_w)$$

其中，x 为萘的转化率；l 为反应管长；T 为反应温度；T_w 为反应管壁温度。

假定反应器入口温度为 340 ℃，萘转化率 $x_0=0$，反应管壁温稳定在 340 ℃，试求当萘转化率为 80% 时，反应器内温度分布和反应管长度。

解 此题是初值问题：

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}l} = 1.1094 \times 10^{11}\left(\dfrac{1-x}{1\,000-0.5x}\right)^{0.38} \exp\left(-\dfrac{14\,098}{T}\right) \\ \dfrac{\mathrm{d}T}{\mathrm{d}l} = 7.6120 \times 10^{12}\left(\dfrac{1-x}{1\,000-0.5x}\right)^{0.38} \exp\left(-\dfrac{14\,098}{T}\right) - 9.9282(T-613) \\ l=0, \quad x_0=0, \quad T_0=613 \end{cases}$$

现取步长 $h=\Delta l=0.01$，用四阶龙格—库塔法进行计算。下面列出部分计算结果：

l/m	0	0.2	0.4	0.5	0.6	0.8	0.9	1.0	1.05
x	0	0.1817	0.3644	0.4492	0.5285	0.6701	0.7322	0.7885	0.8144
T/K	613	618.51	619.14	618.94	618.64	617.89	617.49	617.09	616.89

当转化率 $x=0.80$ 时，约需管长 $l=1.023$ m。由列出的结果可见，床层温度分布是两端低中间高。

二、阿达姆斯法

四阶阿达姆斯预报—校正格式用于求解常微分方程组的初值问题时可表达为

$$\begin{cases} y_{ji+1}^{0} = y_{ji} + \dfrac{h}{24}(55f_{ji} - 59f_{ji-1} + 37f_{ji-2} - 9f_{ji-3}) \\ y_{ji+1} = y_{ji} + \dfrac{h}{24}(9f_{ji+1}^{0} + 19f_{ji} - 5f_{ji-1} + f_{ji-2}) \\ j=1,2,\cdots,m; \quad i=3,4,\cdots,n-1 \end{cases}$$

其中

$$f_{ji}=f_j(x_i,y_{1i},y_{2i},\cdots,y_{mi}),\quad f^0_{ji+1}=f_j(x_{i+1},y^0_{1i+1},y^0_{2i+1},\cdots,y^0_{mi+1});$$

$$h=\frac{1}{n}(x_n-x_0),\quad x_i=x_0+ih\quad(i=0,1,2,\cdots,n)$$

计算时需由经典龙格—库塔法提供开头三步的信息 $y_{j1},y_{j2},y_{j3}(j=1,2,\cdots,m)$。

同样,为了编程方便,令 $y_0=x$,则使原方程组增多一个微分方程

$$\frac{\mathrm{d}y_0}{\mathrm{d}y_0}=\frac{\mathrm{d}y_0}{\mathrm{d}y}=f_0(y_0,y_1,y_2,\cdots,y_m)\equiv 1$$

则有

$$\begin{cases}\dfrac{\mathrm{d}y_j}{\mathrm{d}y_0}=f_j(y_0,y_1,y_2\cdots,y_m)\\ y_j(y_0)=y_{j0}\end{cases}$$

其中,$y_0\in[x_0,x_n]$;$j=0,1,2,\cdots,m$。

对该式施以四阶阿达姆斯预报—校正格式进行计算。

例 8-6 等温下进行连串反应 $A+B\underset{k_2}{\overset{k_1}{\rightleftharpoons}}C\xrightarrow{k_3}D$,其反应动力学模型为

$$\begin{cases}\dfrac{\mathrm{d}c_A}{\mathrm{d}t}=-k_1c_Ac_B+k_2c_C\\ \dfrac{\mathrm{d}c_B}{\mathrm{d}t}=-k_1c_Ac_B+k_2c_C\\ \dfrac{\mathrm{d}c_C}{\mathrm{d}t}=k_1c_Ac_B-(k_2+k_3)c_C\end{cases}$$

已知反应温度下,$k_1=0.13\ \mathrm{L/(mol\cdot min)}$,$k_2=0.049\ 1\ \mathrm{min^{-1}}$,$k_3=0.11\ \mathrm{min^{-1}}$。

反应开始时 $c_{A0}=0.05\ \mathrm{mol/L}$,$c_{B0}=0.1\ \mathrm{mol/L}$,$c_{C0}=c_{D0}=0$。试确定目的产物 C 的最大浓度及最佳反应时间。

解 依题意知本题为初值问题

$$\begin{cases}\dfrac{\mathrm{d}c_A}{\mathrm{d}t}=-k_1c_Ac_B+k_2c_C\\ \dfrac{\mathrm{d}c_B}{\mathrm{d}t}=-k_1c_Ac_B+k_2c_C\\ \dfrac{\mathrm{d}c_C}{\mathrm{d}t}=k_1c_Ac_B-(k_2+k_3)c_C\\ t=0,c_{A0}=0.05,c_{B0}=0.1,c_{C0}=c_{D0}=0\end{cases}$$

取步长 $h=\Delta t=10\ \mathrm{min}$,利用四阶阿达姆斯预报—校正格式进行计算,部分结果为:

t/min	0	10	20	30	50	100
$c_A \times 10^2/(\text{mol} \cdot \text{L}^{-1})$	5.0	4.592 9	4.381 4	4.215 0	3.921 3	3.304 8
$c_B \times 10^2/(\text{mol} \cdot \text{L}^{-1})$	10.0	9.592 9	9.381 4	9.215 0	8.921 3	8.304 8
$c_C \times 10^2/(\text{mol} \cdot \text{L}^{-1})$	0	3.100 4	3.546 3	3.480 6	3.160 4	2.472 4

由计算结果可知,目的的产物 C 的最大浓度为 0.035 46 mol/L,最佳反应时间 $t=20$ min。

第五节　高阶常微分方程的初值问题的数值解法

设有 m 阶常微分方程的初值问题:

$$\begin{cases} \dfrac{\mathrm{d}^m y}{\mathrm{d}x^m} = f\left(x, y, \dfrac{\mathrm{d}y}{\mathrm{d}x}, \dfrac{\mathrm{d}^2 y}{\mathrm{d}x^2}, \cdots, \dfrac{\mathrm{d}^{m-1} y}{\mathrm{d}x^{m-1}}\right), \quad x \in [x_0, x_n] \\ y(x_0) = y_0 \\ \left.\dfrac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=x_0} = y_0 \\ \cdots\cdots \\ \left.\dfrac{\mathrm{d}^{m-1} y}{\mathrm{d}x^{m-1}}\right|_{x=x_0} = y_0^{(m-1)} \end{cases}$$

为求解该初值问题,通常将其转化为等价的一阶常微分方程组。

令

$$y_0 = x, \quad y_1 = y, \quad y_2 = \frac{\mathrm{d}y}{\mathrm{d}x}, \quad \cdots, \quad y_m = \frac{\mathrm{d}^{m-1} y}{\mathrm{d}x^{m-1}}$$

则有

$$\begin{cases} \dfrac{\mathrm{d}y_0}{\mathrm{d}y_0} = \dfrac{\mathrm{d}y_0}{\mathrm{d}x} = 1 \\ \dfrac{\mathrm{d}y_1}{\mathrm{d}y_0} = \dfrac{\mathrm{d}y}{\mathrm{d}x} = y_2 \\ \dfrac{\mathrm{d}y_2}{\mathrm{d}y_0} = \dfrac{\mathrm{d}^2 y}{\mathrm{d}x^2} = y_3 \\ \cdots\cdots \\ \dfrac{\mathrm{d}y_{m-1}}{\mathrm{d}y_0} = \dfrac{\mathrm{d}^{m-1} y}{\mathrm{d}x^{m-1}} = y_m \\ \dfrac{\mathrm{d}y_m}{\mathrm{d}y_0} = \dfrac{\mathrm{d}^m y}{\mathrm{d}x^m} = f(y_0, y_1, \cdots, y_m) \\ y_j(x_0) = y_0^{(j-1)} \quad (j = 0, 1, 2, \cdots, m) \end{cases}$$

其中 $y_0(x_0) = y_0^{-1} = x_0$。

这样，便将 m 阶的常微分方程的初值问题转化为 $m+1$ 个方程所构成的常微分方程组的初值问题，即可利用前面介绍的方法求解。

例 8-7 求解二阶初值问题

$$\begin{cases}\dfrac{d^2y}{dx^2}=\dfrac{2x}{1+x^2}\dfrac{dy}{dx}, & x\in[0,1]\\ y(0)=1, \quad y'(0)=3\end{cases}$$

解 令

$$y_0=x, \quad y_1=y, \quad y_2=\frac{dy}{dy_0}=\frac{dy}{dx}=\frac{dy_1}{dy_0}$$

则

$$\begin{cases}\dfrac{dy_2}{dy_0}=\dfrac{2y_0}{1+y_0^2}y_2, & y_0\in[0,1]\\ \dfrac{dy_1}{dy_0}=y_2\\ y_0(0)=x_0=0, \quad y_1(0)=1, \quad y_2(0)=3\end{cases}$$

取步长 $h=0.1$，利用经典龙格—库塔法计算得：

x_i	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
y_i	1	1.301	1.608	1.927	2.264	2.642	3.015	3.442	3.911	4.428	4.999

第六节 常微分方程边值问题的数值解法

常微分方程的边值问题是指给出了自变量在某一区间两端点处的函数值的情形，也称为两边值问题。现以二阶常微分方程的边值问题为例进行讨论。

一般二阶常微分方程的边值问题可表示为

$$\frac{d^2y}{dx^2}=f\left(x,y,\frac{dy}{dx}\right), \quad x\in[x_0,x_n]$$

其边界条件有以下 3 类。

第一边界条件：$y(x_0)=y_0, \quad y(x_n)=y_n$；

第二边界条件：$\left.\dfrac{dy}{dx}\right|_{x=x_0}=y_0', \quad \left.\dfrac{dy}{dx}\right|_{x=x_n}=y_n'$；

第三边界条件：$\left.\dfrac{dy}{dx}\right|_{x=x_0}-\alpha y(x_0)=a, \quad \left.\dfrac{dy}{dx}\right|_{x=x_n}+\beta y(x_n)=b$。其中，$\alpha\geqslant 0, \alpha+\beta>0$。

边值问题有各种不同的数值解法，常用的有打靶法、差分法和正交配置。下面分别讨论之。

一、解边值问题的打靶法

打靶法就是将边值问题转化为初值问题，通过逐次逼近法来求解。其具体做法是：先不考虑右端边界条件，而是在左端再设定一个条件，从而使边界问题转化为初值问题，然后，就可利用各种解初值问题的数值方法进行求解。

对于第一边值条件的二阶方程

$$\frac{d^2y}{dx^2}=f\left(x,y,\frac{dy}{dx}\right),\quad x\in[x_0,x_n]$$

将其转化为初值问题

$$\begin{cases}\dfrac{d^2y}{dx^2}=f\left(x,y,\dfrac{dy}{dx}\right),\quad x\in[x_0,x_n]\\ y(x_0)=y_0\\ \left.\dfrac{dy}{dx}\right|_{x=x_0}=m\end{cases}$$

其中，m 是人为的设定值。仅当 $m=m^*$，且满足 $y(x_n,m^*)=y_n$ 时，所得 $y(x)$ 便为所求的解。因此，问题的实质是选择 m 值，相当于求如下非线性方程的根：

$$f(m)=y(x_n,m)-y_n=0$$

这可利用非线性方程求解法中的非导数法求解，如弦位法等。

显然，打靶法是一种逐次逼近法，即先根据实际对客观过程的理解设定一个 m 值。通过解初值问题

$$\begin{cases}\dfrac{d^2y}{dx^2}=f\left(x,y,\dfrac{dy}{dx}\right)\\ y(x_0)=y_0\\ \left.\dfrac{dy}{dx}\right|_{x=x_0}=m\end{cases}$$

求得一个 $y(x_n,m)$，若有 $|y(x_n,m)-y_n|\leqslant\varepsilon$ 成立，则 $y(x_n,m)$ 便为所求的解。否则，重新设定 m 直到满足精度要求。

其几何意义如图 8-3 所示，即要从微分方程 $\dfrac{d^2y}{dx^2}=f\left(x,y,\dfrac{dy}{dx}\right)$ 的经过点 $A(x_0,y_0)$ 出发，从具有不同斜率的一族积分曲线中找一条通过点 $B(x_n,y_n)$ 的曲线。由于它和弹道问题类似，故称为打靶法，也即是寻找曲线初始的适宜的斜率，使其射中目标 $y(x_n,m)=y_n$。这样，$y(x_n,m)$ 是 m 的函数，于是问题就变为求取方程 $y(x_n,m)-y_n=0$ 的根的问题。若采用弦位法求该方程的根，则

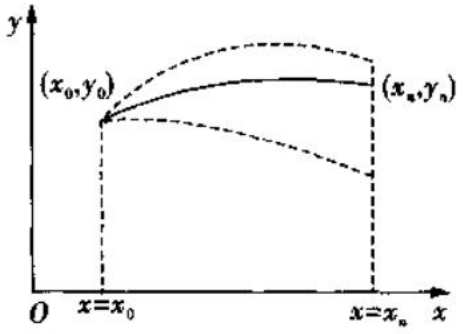


图 8-3

$$m_{k+1}=m_k-\frac{m_k-m_{k-1}}{y(x_n,m_k)-y(x_n,m_{k-1})}y(x_n,m_k)\quad(k=1,2,\cdots)$$

于是,打靶法的计算步骤为:

(1)设定初始值 m_0 和 m_1,给定精度 δ 或 ε。

(2)应用龙格—库塔法或阿达姆斯法求解初值问题

$$\begin{cases}\dfrac{d^2y}{dx^2}=f\left(x,y,\dfrac{dy}{dx}\right)\\ y(x_0)=y_0\\ y'(x_0)=m\end{cases}$$

得 $y(x_n,m_0)$ 和 $y(x_n,m_1)$。

(3)利用弦位法求 $m_{k+1}(k=1,2,\cdots)$。

(4)若 $|m_{k+1}-m_k|\leqslant\delta$ 或 $|y(x_n,m_{k+1})-y_n|\leqslant\varepsilon$,则初值问题的解 $y(x,m_{k+1})$ 便为边值问题的解;否则,利用 m_{k+1} 回第(2)步重新计算。

例 8-8 应用打靶法求解边值问题

$$\begin{cases}\dfrac{d^2y}{dx^2}=8-\dfrac{y}{4},\quad x\in[0,10]\\ y(0)=0\\ y(10)=0\end{cases}$$

解 首先把边值问题转变为初值问题

$$\begin{cases}\dfrac{d^2y}{dx^2}=8-\dfrac{y}{4},\quad x\in[0,10]\\ y(0)=0,\quad y'(0)=m\end{cases}$$

再把初值问题化为等价的方程组的初值问题,即令 $y_0=x,y_1=y,y_2=\dfrac{dy}{dx}$,于是

$$\begin{cases}\dfrac{dy_1}{dy_0}=y_2\\ \dfrac{dy_2}{dy_0}=8-\dfrac{y_1}{4},\qquad y_0\in[0,10]\\ y_0(0)=x_0=0\\ y_1(0)=0\\ y_2(0)=m\end{cases}$$

取 $m_0=10,m_1=11$,应用弦位法迭代求解。此题的解析解为

$$y=23.9047\sin\frac{x}{2}-32\cos\frac{x}{2}+32$$

为便于比较,现将部分计算结果列表如下:

x_i	0	1	2	3	4	5
打靶法 y_i	0	15.377 9	34.825 4	53.581 2	67.053 1	71.942 9
解析值 y_i	0	15.377 9	34.825 4	53.581 2	67.053 2	71.942 9

x_i	6	7	8	9	10
打靶法 y_i	67.053 2	53.581 3	34.825 6	15.378 0	0.000 01
解析值 y_i	67.053 2	53.581 2	34.825 5	15.377 9	0

计算得 $m^* = 11.952\ 35$。计算结果表明打靶法有很高的精度。

二、解边值问题的差分法

差分法的基本思想是利用差商代替微商，使微分方程离散化，变为代数方程组，然后再应用直接法解得到的代数方程组，因此，也称其为矩阵法。

设有线性二阶常微分方程的边值问题

$$\begin{cases}\dfrac{d^2y}{dx^2} = P(x)\dfrac{dy}{dx} + q(x)y + r(x), \quad x \in [x_0, x_n] \\ y(x_0) = y_0, \quad y(x_n) = y_n\end{cases}$$

首先，将区间$[x_0, x_n]$ n 等分，则步长 $h = \dfrac{1}{n}(x_n - x_0)$，分点 $x_i = x_0 + ih\ (i = 0, 1, 2, \cdots, n)$。然后，在每个节点上，用二阶中心差商和一阶中心差商分别代替二阶和一阶微商，即

$$\frac{d^2y}{dx^2} = \frac{y(x_{i+1}) - 2y(x_i) + y(x_{i-1})}{h^2}$$

$$\frac{dy}{dx} = \frac{y(x_{i+1}) - y(x_{i-1})}{2h}$$

取 $y_{i-1} \approx y(x_{i-1}), y(x_i) \approx y_i, y_{i+1} \approx y(x_{i+1})$ 代入微分方程整理得

$$\left[1 + \frac{h}{2}P(x_i)\right]y_{i-1} - [2 + h^2 q(x_i)]y_i + \left[1 - \frac{h}{2}P(x_i)\right]y_{i+1} = h^2 r(x_i)$$

$$(i = 1, 2, \cdots, n-1)$$

对于第一边值问题

$$y_{i-1} = y_0 = y(x_0) \quad (i = 1)$$

$$y_{i+1} = y_n = y(x_n) \quad (i = n-1)$$

于是，可得三对角线性方程组

$$\begin{cases} -[2+h^2q(x_1)]y_1+\left[1-\dfrac{h}{2}P(x_1)\right]y_2=h^2r(x_1)-\left[1+\dfrac{h}{2}P(x_1)\right]y_0 \\ \left[1+\dfrac{h}{2}P(x_2)\right]y_1-[2+h^2q(x_2)]y_2+\left[1-\dfrac{h}{2}P(x_2)\right]y_3=h^2r(x_2) \\ \cdots\cdots \\ \left[1+\dfrac{h}{2}P(x_{n-2})\right]y_{n-3}-[2+h^2q(x_{n-2})]y_{n-2}+\left[1-\dfrac{h}{2}P(x_{n-2})\right]y_{n-1}=h^2r(x_{n-2}) \\ \left[1+\dfrac{h}{2}P(x_{n-1})\right]y_{n-2}-[2+h^2q(x_{n-1})]y_{n-1}=h^2r(x_{n-1})-\left[1-\dfrac{h}{2}P(x_{n-1})\right]y_n \end{cases}$$

若令

$$a_i=\left[1+\frac{h}{2}P(x_i)\right],\quad b_i=-[2+h^2q(x_i)],\quad c_i=\left[1-\frac{h}{2}P(x_i)\right],\quad d_i=h^2r(x_i)$$

其中，$i=1,2,\cdots,n-1$。则三对角线性方程组可写为

$$\begin{bmatrix} b_1 & c_1 & & & \\ a_2 & b_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & a_{n-2} & b_{n-2} & c_{n-2} \\ & & & a_{n-1} & b_{n-1} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-2} \\ y_{n-1} \end{bmatrix} = \begin{bmatrix} d_1-a_1y_0 \\ d_2 \\ \vdots \\ d_{n-2} \\ d_{n-1}-c_{n-1}y_n \end{bmatrix}$$

若取 $P=\max\limits_{x_0\leqslant x\leqslant x_n}|P(x)|$，$h\leqslant\dfrac{2}{P}$，则可保证主对角线占优。该三对角方程组有唯一解，可用追赶法求解。

综上所述，差分法计算步骤为：

(1)计算 $P=\max\limits_{x_0\leqslant x\leqslant x_n}|P(x)|$；

(2)取 $h\leqslant\dfrac{2}{P}$，$n=(x_n-x_0)/h$；

(3)对于 $i=1,2,\cdots,n-1$，计算

$$x_i=x_0+ih,$$

$$a_i=1+\frac{1}{2}hP(x_i),\quad b_i=-[2+h^2q(x_i)],\quad c_i=2-a_i,\quad d_i=h^2r(x_i);$$

(4)令 $d_1=d_1-a_1y_0$，$d_{n-1}=d_{n-1}-c_{n-1}y_n$；

(5)由追赶法求 $y_1,y_2,\cdots,y_{n-1}$；

(6)输出 $x_i,y_i(i=0,1,2,\cdots,n)$。

例 8-9 翅片长度 $L=0.305$ m 的矩形翅片散热器的微分方程为

$$\begin{cases} \dfrac{d^2T}{dl^2}=2.35(T-37.8),\quad l\in[0,0.305] \\ T(0)=93.3,\quad T(0.305)=65.6 \end{cases}$$

试用差分法计算翅片内的温度分布。

解 将区间[0,0.305]5 等分，则 $h=\Delta l=\frac{0.305}{5}=0.061$，这里，$P(l)=0$，$q(l)=2.35$，$r(l)=-88.83$，则得三对角方程组

$$\begin{bmatrix} -2.008\ 74 & 1.0 & 0 & 0 \\ 1.0 & -2.008\ 74 & 1.0 & 0 \\ 0 & 1.0 & -2.008\ 74 & 1.0 \\ 0 & 0 & 1.0 & -2.008\ 74 \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{bmatrix} = \begin{bmatrix} -93.63 \\ -0.33 \\ -0.33 \\ -65.93 \end{bmatrix}$$

解得翅片内的温度分布为：

l/m	0	0.061	0.122	0.183	0.244	0.305
T/℃	93.30	87.02	81.14	75.63	70.45	65.60

三、解边值问题的正交配置法

正交配置法是近年来在化学工程计算中得到广泛应用的一种数值方法。利用它既可求得微分方程的解函数在各配置点的值，也可求得微分方程的近似解析解。

正交配置法的基本思想是利用正交多项式 $P_i(x)$ 作为试解的基函数，即

$$y(x)=y_0+(x-x_0)\sum_{i=1}^{n}a_iP_i(x)$$

其中，a_i 为拟合系数；$P_i(x)=\sum_{j=0}^{i}c_jx^j$，为 i 次正交多项式。

该式中的各系数 c_j 应满足能使 $P_1(x)$ 与 $P_0(x)$ 正交，$P_2(x)$ 与 $P_1(x)$、$P_0(x)$ 正交，P_i 与各 $P_k(x)(k\leqslant i-1)$ 正交。即

$$\int_0^1 P_i(x)P_k(x)\mathrm{d}x=0$$

若假设：$P_0(x)=1$，$P_1=1+c_1x$，$P_2=1+bx+x_2x^2$。$P_1(x)$ 可由下式确定

$$\int_0^1 P_0(x)P_1(x)\mathrm{d}x=\int_0^1(1+c_1x)\mathrm{d}x=0$$

求得 $c_1=-2$。

则

$$P_1(x)=1-2x$$

$P_2(x)$ 可由以下两式确定：

$$\begin{cases}\int_0^1 P_0(x)P_2(x)\mathrm{d}x = \int_0^1 (1 + bx + c_2x^2)\mathrm{d}x = 0 \\ \int_0^1 P_1(x)P_2(x)\mathrm{d}x = \int_0^1 (1 - 2x)(1 + bx + c_2x^2)\mathrm{d}x = 0\end{cases}$$

求得 $b = -6, c_2 = 6$，即

$$P_2(x) = 1 - 6x + 6x^2$$

由 $P_1(x) = 0$，可得 $x = \frac{1}{2}$；由 $P_2(x) = 0$，可得 $x = \frac{1}{2}\left[1 \pm \frac{\sqrt{3}}{3}\right]$。它们即为内配置点数为 1 和 2 时内配置点的位置。当需要的内配置点数为 n 时，可用同样的方法求得 n 次正交多项式 $P_n(x)$，它在区间[0,1]内有 n 个根，这些根即为配置点的位置。由此可见，在正交配置法中，试解函数的形式和配置点的位置都是确定的。

现将正交配置法用于求解常微分方程的两点边值问题。设常微分方程的边值问题

$$\begin{cases}\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = f\left(x, y, \frac{\mathrm{d}y}{\mathrm{d}x}\right), \quad x \in [0,1] \\ y(0) = 0, \quad y(1) = 1\end{cases}$$

对齐次边界条件，都可通过无因次化写成上述标准形式。

为满足上述边界条件，取试解函数

$$y = x + x(1 - x)\sum_{i=1}^{n} a_iP_{i-1}(x)$$

上式可改写为

$$y = \sum_{i=1}^{n+2} b_iP_{i-1}(x)$$

为了简化导数矩阵的推导，其也可写为

$$y = \sum_{i=1}^{n+2} d_ix^{i-1}$$

由该式求得 y 的一阶、二阶导数后，可在各配置点计算 y，$\frac{\mathrm{d}y}{\mathrm{d}x}$ 和 $\frac{\mathrm{d}^2y}{\mathrm{d}x^2}$ 的值。

即

$$y(x_j) = \sum_{i=1}^{n+2} d_ix_j^{i-1}$$

$$\frac{\mathrm{d}y(x_j)}{\mathrm{d}x} = \sum_{i=1}^{n+2} d_i(i-1)x_j^{i-2}$$

$$\frac{\mathrm{d}^2y(x_j)}{\mathrm{d}x^2} = \sum_{i=1}^{n+2} d_i(i-1)(i-2)x_j^{i-3}$$

将它们写为矩阵形式为

$$Y = Qd$$

$$\frac{dY}{dx} = Cd$$

$$\frac{d^2Y}{dx^2} = Dd$$

其中，矩阵 Q、C、D 都是 $n+2$ 阶方阵，其元素分别为

$$q_{ji} = x_j^{i-1}$$

$$c_{ji} = (i-1)x_j^{i-2}$$

$$d_{ji} = (i-1)(i-2)x_j^{i-3}$$

由于

$$d = Q^{-1}Y$$

则

$$\frac{dY}{dx} = Cd = CQ^{-1}Y \approx AY$$

$$\frac{d^2Y}{dx^2} = Dd = DQ^{-1}Y = BY$$

可见，在任一配置点的导数值均可用所有配置点上的函数值计算。

将各配置点 $x_j(j=1,2,\cdots,n+2)$ 上的函数值 $y(x_j)$ 的一阶导数值 $\frac{dy(x_j)}{dx}$，二阶导数值 $\frac{d^2y(x_j)}{dx^2}$ 代入原微分方程，可得到 $n+2$ 个代数方程

$$\sum_{i=1}^{n+2} b_{ji}y(x_j) = f\left[x_j, y(x_j), \sum_{i=1}^{n+2} \alpha_{ji}y(x_j)\right] \quad (j = 1,2,\cdots,n+2)$$

结合边界条件，由该方程组可解得各配置点上的函数值 $y(x_j)$。若将求得的函数值 $y(x_j)$ $(j=1,2,\cdots,n+2)$ 代入 $d = Q^{-1}Y$，解得 $d_i(i=1,2,\cdots,n+2)$，即得到

$$y = \sum_{i=1}^{n+2} d_i x^{i-1}$$

它为微分方程的近似解析解。

例 8-10 存在轴向分散的绝热管式反应器中进行一组反应时，无因次物料衡算方程和能量衡算方程为

$$\frac{1}{p_{em}}\frac{d^2x_A}{d\xi^2} - \frac{dx_A}{d\xi} + D_a \exp\left[\varepsilon\left(1 - \frac{1}{\theta}\right)\right](1 - x_A) = 0$$

和

$$\frac{1}{p_{eh}}\frac{d^2\theta}{d\xi^2} - \frac{d\theta}{d\xi} + \beta D_a \exp\left[\varepsilon\left(1 - \frac{1}{\theta}\right)\right](1 - x_A) = 0$$

边界条件为

$$\xi=0,\quad x_A-\frac{1}{p_{em}}\frac{\mathrm{d}x_A}{\mathrm{d}\xi}=0,\quad 1-\theta+\frac{1}{p_{eh}}\frac{\mathrm{d}\theta}{\mathrm{d}\xi}=0;$$

$$\xi=1,\quad \frac{\mathrm{d}x_A}{\mathrm{d}\xi}=\frac{\mathrm{d}\theta}{\mathrm{d}\xi}=0$$

用正交配置法计算当 $p_{em}=p_{eh}=5$，$D_a=0.15$，$\varepsilon=17$，$\beta=0.35$ 时，反应器出口处的转化率 x_A 和无因次温度 θ。

解 取内配置点数为3，根据正交多项式的性质，可求得配置点位置为

$$\xi_1=0,\quad \xi_2=0.112\,7,\quad \xi_3=0.500\,0,\quad \xi_4=0.887\,3,\quad \xi_5=1.000$$

于是由 $q_{ji}=x_j^{i-1}$，$c_{ij}=(i-1)x_j^{i-2}$，$d_{ji}=(i-1)(i-2)x_j^{i-3}$ 得

$$Q=[\xi_j^{i-1}]=\begin{bmatrix}1.000\,0 & 0.000\,0 & 0.000\,0 & 0.000\,0 & 0.000\,0\\ 1.000\,0 & 0.112\,7 & 0.012\,7 & 0.001\,431 & 0.000\,161\,3\\ 1.000\,0 & 0.500\,0 & 0.250\,0 & 0.125\,0 & 0.062\,50\\ 1.000\,0 & 0.887\,3 & 0.787\,3 & 0.698\,6 & 0.619\,8\\ 1.000\,0 & 1.000\,0 & 1.000\,0 & 1.000\,0 & 1.000\,0\end{bmatrix}$$

$$C=[(i-1)\xi_j^{i-2}]=\begin{bmatrix}0.000\,0 & 1.000\,0 & 0.000\,0 & 0.000\,0 & 0.000\,0\\ 0.000\,0 & 1.000\,0 & 0.225\,4 & 0.038\,10 & 0.005\,726\\ 0.000\,0 & 1.000\,0 & 1.000\,0 & 0.750\,0 & 0.500\,0\\ 0.000\,0 & 1.000\,0 & 1.774\,6 & 2.361\,9 & 2.794\,3\\ 0.000\,0 & 1.000\,0 & 2.000\,0 & 3.000\,0 & 4.000\,0\end{bmatrix}$$

$$D=[(i-1)(i-2)\xi_j^{i-3}]=\begin{bmatrix}0.000\,0 & 0.000\,0 & 2.000\,0 & 0.000\,0 & 0.000\,0\\ 0.000\,0 & 0.000\,0 & 2.000\,0 & 0.676\,2 & 0.152\,4\\ 0.000\,0 & 0.000\,0 & 2.000\,0 & 3.000\,0 & 3.000\,0\\ 0.000\,0 & 0.000\,0 & 2.000\,0 & 5.323\,8 & 9.447\,6\\ 0.000\,0 & 0.000\,0 & 2.000\,0 & 6.000\,0 & 12.000\,0\end{bmatrix}$$

对矩阵 Q 求逆得

$$Q^{-1}=\begin{bmatrix}1.000\,0 & 0.000\,0 & 0.000\,0 & 0.000\,0 & 0.000\,0\\ -13.000\,0 & 14.788\,3 & -2.666\,7 & 1.878\,4 & -1.000\,0\\ 42.000\,0 & -61.031\,6 & 29.333\,3 & -22.301\,7 & 12.000\,0\\ -50.000\,0 & 79.576\,6 & -53.333\,3 & 53.756\,7 & 30.000\,0\\ 20.000\,0 & -33.333\,3 & 26.666\,7 & -33.333\,3 & 20.000\,0\end{bmatrix}$$

再由 $A=CQ^{-1}$ 和 $B=DQ^{-1}$ 得

$$A=\begin{bmatrix}-13.000\,0 & 14.788\,3 & -2.666\,7 & 1.878\,4 & -1.000\,0\\ -5.323\,8 & 3.873\,0 & 2.065\,6 & -1.291\,0 & 0.676\,2\\ 1.500\,0 & -3.227\,5 & 0.000\,0 & 3.227\,5 & -1.500\,0\\ -0.676\,3 & 1.291\,0 & -2.065\,6 & -3.873\,0 & 5.323\,8\\ 1.000\,0 & -1.878\,4 & 2.666\,7 & -14.788\,3 & 13.000\,0\end{bmatrix}$$

$$B=\begin{bmatrix} 84.000\,0 & -122.063\,2 & 58.666\,7 & -44.603\,5 & 24.000\,0 \\ 53.237\,9 & -73.333\,3 & 26.666\,7 & -13.333\,3 & 6.762\,1 \\ -6.000\,0 & 16.666\,7 & -21.333\,3 & 16.666\,7 & -6.000\,0 \\ 6.762\,1 & -13.333\,3 & 26.666\,7 & -73.333\,3 & 53.237\,9 \\ 24.000\,0 & -44.603\,5 & 58.666\,7 & -122.063\,2 & 84.000\,0 \end{bmatrix}$$

将$\frac{\mathrm{d}x_A}{\mathrm{d}\xi}\approx Ax_A,\frac{\mathrm{d}^2x_A}{\mathrm{d}\xi^2}=Bx_A,\frac{\mathrm{d}\theta}{\mathrm{d}\xi}=A\theta,\frac{\mathrm{d}^2\theta}{\mathrm{d}\xi^2}=B\theta$代入微分方程和边界条件得代数方程组

$$\begin{cases} 3.6x_{A1}-2.957\,7x_{A2}+0.533\,3x_{A3}-0.375\,7x_{A4}+0.2x_{A5}=0 \\ 15.971\,4x_{A1}-18.539\,7x_{A2}+3.267\,7x_{A3}-1.375\,7x_{A4}+0.676\,2x_{A5}+ \\ \quad 0.15\exp\left[17\left(1-\frac{1}{\theta_2}\right)\right](1-x_{A2})=0 \\ -2.7x_{A1}+6.560\,9x_{A2}-4.266\,7x_{A3}+0.105\,8x_{A4}+0.3x_{A5}+ \\ \quad 0.15\exp\left[17\left(1-\frac{1}{\theta_3}\right)\right](1-x_{A3})=0 \\ 2.028\,6x_{A1}-3.957\,7x_{A2}+7.398\,9x_{A3}-10.793\,7x_{A4}+5.323\,8x_{A5}+ \\ \quad 0.15\exp\left[17\left(1-\frac{1}{\theta_4}\right)\right](1-x_{A4})=0 \\ x_{A1}-1.878\,4x_{A2}+2.666\,7x_{A3}-14.788\,3x_{A4}+13x_{A5}=0 \\ 1-3.6\theta_1+2.957\,7\theta_2-0.533\,3\theta_3+0.375\,7\theta_4-0.2\theta_5=0 \\ 15.971\,4\theta_1-18.539\,7\theta_2+3.267\,7\theta_3-1.375\,7\theta_4+0.676\,2\theta_5+ \\ \quad 0.052\,5\exp\left[17\left(1-\frac{1}{\theta_2}\right)\right](1-x_{A2})=0 \\ -2.7\theta_1+6.560\,9\theta_2-4.266\,7\theta_3+0.105\,8\theta_4+0.3\theta_5+ \\ \quad 0.052\,5\exp\left[17\left(1-\frac{1}{\theta_3}\right)\right](1-x_{A3})=0 \\ 2.028\,6\theta_1-3.957\,7\theta_2+7.398\,9\theta_3-10.793\,7\theta_4+5.323\,8\theta_5+ \\ \quad 0.052\,5\exp\left[17\left(1-\frac{1}{\theta_4}\right)\right](1-x_{A4})=0 \\ \theta_1-1.878\,4\theta_2+2.666\,7\theta_3-14.788\,3\theta_4+13\theta_5=0 \end{cases}$$

假定初值$x_{Ai}=0(i=1,2,\cdots,5),\theta_i=1(i=1,2,\cdots,5)$，进行迭代求解，经17次迭代求得

$$x_{A1}=0.062\,7,\quad x_{A2}=0.095\,4,\quad x_{A3}=0.271\,6,\quad x_{A4}=0.532\,9,\quad x_{A5}=0.532\,9;$$
$$\theta_1=1.021\,9,\quad \theta_2=1.033\,4,\quad \theta_3=1.095\,0,\quad \theta_4=1.186\,5,\quad \theta_5=1.195\,8$$

习 题

1. 用显式尤拉法解初值问题(取 $h=0.1$):

$$\begin{cases}\dfrac{dy}{dx}+2xy=4, & x\in[0,0.5]\\ y(0)=0\end{cases}$$

2. 用改进尤拉法求解下列初值问题:

(1) $\begin{cases}\dfrac{dy}{dx}=\dfrac{3x^2}{2y}, & x\in[0,2]\\ y(0)=2\end{cases}$ 取 $h=0.4$。

(2) $\begin{cases}\dfrac{dy}{dx}=\sqrt{x+y}, & x\in[0,1]\\ y(0)=0.36\end{cases}$ 取 $h=0.2$。

3. 已知可逆反应 $A\underset{k_2}{\overset{k_1}{\rightleftharpoons}}B$ 的速率方程为 $\dfrac{dc_A}{dt}=k_2c_B-k_1c_A$,反应温度下 $k_1=10,k_2=5$,原料中 $c_{A0}=1.0\ \text{mol/L},c_{B0}=0.0\ \text{mol/L}$,现取 $h=0.05$,用龙格—库塔法计算 $c_A=0.1$ 和 $c_B=0.1$ 时所需的反应时间。

4. 试用四阶阿达姆斯预报—校正格式计算初值问题(取 $h=0.2,\varepsilon=10^{-4}$):

$$\begin{cases}\dfrac{dy}{dx}=x-y^2, & x\in[0,1]\\ y(0)=1\end{cases}$$

5. 利用改进尤拉法解初值问题(取 $h=0.5$):

$$\begin{cases}\dfrac{d^2y}{dx^2}=xy, & x\in[0,1]\\ y(0)=0\\ y'(0)=1\end{cases}$$

6. 用龙格—库塔法解初值问题(取 $h=0.2$):

$$\begin{cases}\dfrac{d^2y}{dx^2}=\dfrac{x^2-y^2}{1+\left(\dfrac{dy}{dx}\right)^2}, & x\in[0,1.5]\\ y(0)=1\\ y'(0)=0\end{cases}$$

7. 有如下反应:

$$A\xrightarrow{k_1}B$$

$$B+C\xrightarrow{k_2}A+C$$

$$2B \xrightarrow{k_3} C + B$$

其动力学方程为

$$\begin{cases} \dfrac{dc_A}{dt} = -k_1 c_A + k_2 c_B c_C \\ \dfrac{dc_B}{dt} = k_1 c_A - k_2 c_B c_C - k_3 c_B^2 \\ \dfrac{dc_C}{dt} = k_3 c_B^2 \\ c_A(0) = 1, \quad c_B(0) = c_C(0) = 0 \end{cases}$$

其中，$k_1 = 0.08$，$k_2 = 2 \times 10^4$，$k_3 = 6 \times 10^7$。试用阿达姆斯预报—校正格式进行求解。

8. 试用打靶法计算：

$$\begin{cases} \dfrac{d^2y}{dx^2} = 2x - 4y + 1, \quad x \in [0,1] \\ y(0) = 0 \\ y(1) = 1 \end{cases}$$

9. 用差分法计算：

$$\begin{cases} \dfrac{d^2y}{dx^2} = 8 - \dfrac{1}{4}y, \quad x \in [0,10] \\ y(0) = 0, \quad y(10) = 0 \end{cases}$$

参考文献

1 易大义,蒋淑豪,李有法. 数值方法. 杭州:浙江科学技术出版社,1984
2 W. E. 密伦. 数值计算. 北京:科学出版社,1959
3 张吉瑞. 化工数值方法. 北京:中国石化出版社,1995
4 周爱月. 化工数学. 北京:化学工业出版社,1995
5 王树森. 化学工程计算方法. 北京:化学工业出版社,1989
6 袁谓康,朱开宏. 化学反应工程分析. 上海:华东理工大学出版社,1995
7 [美]B. A 芬莱逊著. 李绍芬等译. 化工非线性分析. 天津:天津大学出版社,1992
8 江体乾. 化工数据处理. 北京:化学工业出版社,1984